Encyclopedia of Energy Engineering *and* Technology

Edited by

Barney L. Capehart

CRC Press
Taylor & Francis Group

Volume 1

Encyclopedia of Energy Engineering *and* Technology

Encyclopedias from Taylor & Francis Group

Agriculture Titles

Dekker Agropedia Collection (Eight Volume Set)
ISBN: 0-8247-2194-2 13-Digit ISBN: 978-0-8247-2194-7

Encyclopedia of Agricultural, Food, and Biological Engineering
Edited by Dennis R. Heldman
ISBN: 0-8247-0938-1 13-Digit ISBN: 978-0-8247-0938-9

Encyclopedia of Animal Science
Edited by Wilson G. Pond and Alan Bell
ISBN: 0-8247-5496-4 13-Digit ISBN: 978-0-8247-5496-9

Encyclopedia of Pest Management
Edited by David Pimentel
ISBN: 0-8247-0632-3 13-Digit ISBN: 978-0-8247-0632-6

Encyclopedia of Pest Management, Volume II
Edited by David Pimentel
ISBN: 1-4200-5361-2 13-Digit ISBN: 978-1-4200-5361-6

Encyclopedia of Plant and Crop Science
Edited by Robert M. Goodman
ISBN: 0-8247-0944-6 13-Digit ISBN: 978-0-8247-0944-0

Encyclopedia of Soil Science, Second Edition (Two Volume Set)
Edited by Rattan Lal
ISBN: 0-8493-3830-1 13-Digit ISBN: 978-0-8493-3830-4

Encyclopedia of Water Science
Edited by B.A. Stewart and Terry Howell
ISBN: 0-8247-0948-9 13-Digit ISBN: 978-0-8247-0948-8

Chemistry Titles

Encyclopedia of Chromatography, Second Edition (Two Volume Set)
Edited by Jack Cazes
ISBN: 0-8247-2785-1 13-Digit ISBN: 978-0-8247-2785-7

Encyclopedia of Surface and Colloid Science, Second Edition (Eight Volume Set)
Edited by P. Somasundaran
ISBN: 0-8493-9615-8 13-Digit ISBN: 978-0-8493-9615-1

Encyclopedia of Supramolecular Chemistry (Two Volume Set)
Edited by Jerry L. Atwood and Jonathan W. Steed
ISBN: 0-8247-5056-X 13-Digit ISBN: 978-0-8247-5056-5

Engineering Titles

Encyclopedia of Chemical Processing (Five Volume Set)
Edited by Sunggyu Lee
ISBN: 0-8247-5563-4 13-Digit ISBN: 978-0-8247-5563-8

Encyclopedia of Corrosion Technology, Second Edition
Edited by Philip A. Schweitzer, P.E.
ISBN: 0-8247-4878-6 13-Digit ISBN: 978-0-8247-4878-4

Dekker Encyclopedia of Nanoscience and Nanotechnology (Five Volume Set)
Edited by James A. Schwarz, Cristian I. Contescu, and Karol Putyera
ISBN: 0-8247-5055-1 13-Digit ISBN: 978-0-8247-5055-8

Encyclopedia of Optical Engineering (Three Volume Set)
Edited by Ronald G. Driggers
ISBN: 0-8247-0940-3 13-Digit ISBN: 978-0-8247-0940-2

Business Titles

Encyclopedia of Library and Information Science, Second Edition (Four Volume Set)
Edited by Miriam Drake
ISBN: 0-8247-2075-X 13-Digit ISBN: 978-0-8247-2075-9

Encyclopedia of Public Administration and Public Policy (Two Volume Set)
Edited by Jack Rabin
ISBN: 0-8247-4748-8 13-Digit ISBN: 978-0-8247-4748-0

Coming Soon

Encyclopedia of Public Administration and Public Policy, Second Edition
Edited by Jack Rabin
ISBN: 1-4200-5275-6 13-Digit ISBN: 978-1-4200-5275-6

Encyclopedia of Water Science, Second Edition (Two Volume Set)
Edited by Stanley W. Trimble
ISBN: 0-8493-9627-1 13-Digit ISBN: 978-0-8493-9627-4

Encyclopedia of Wireless and Mobile Communications (Three Volume Set)
Edited by Borko Furht
ISBN: 1-4200-4326-9 13-Digit ISBN: 978-1-4200-4326-6

These titles are available both in print and online. To order, visit:
www.crcpress.com
Telephone: 1-800-272-7737
Fax: 1-800-374-3401
E-Mail: orders@crcpress.com

Volume 1, 2 and 3

Encyclopedia of Energy Engineering *and* Technology

Edited by

Barney L. Capehart

University of Florida
Gainesville, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an informa business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-0-8493-3653-9 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Encyclopedia of energy engineering and technology / edited by Barney L. Capehart.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-8493-3653-9 (alk. paper : set)
ISBN-10: 0-8493-3653-8 (alk. paper : set)
ISBN-13: 978-0-8493-5039-9 (alk. paper : v. 1)
ISBN-10: 0-8493-5039-5 (alk. paper : v. 1)
[etc.]
1. Power resources--Encyclopedias. 2. Power (Mechanics)--Encyclopedias. I. Capehart, B. L. (Barney L.)

TJ163.2.E385 2007
621.04203--dc22 2007007878

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

In memory of my mother and father
and
for my grandchildren Hannah and Easton.
May their energy future be efficient and sustainable.

Editorial Advisory Board

Dr. Stephen A. Roosa, CEM, CIAQP, CMVP, CBEP, CDSM, MBA
Performance Engineer, Energy Systems Group, Inc., Louisville, Kentucky, U.S.A.

Dr. Wayne H. Smith
Interim Dean for Research and Interim Director, Florida Agricultural Experiment Station, University of Florida, Gainesville, Florida, U.S.A.

Prof. James S. Tulenko
Laboratory for Development of Advanced Nuclear Fuels and Materials, University of Florida, Gainesville, Florida, U.S.A.

Dr. Dan Turner
Director, Energy Systems Laboratory, Texas A&M University, College Station, Texas, U.S.A.

Dr. Wayne C. Turner
Editor, Energy Management Handbook, and Regents Professor, Industrial Engineering and Management, Oklahoma State University, Stillwater, Oklahoma, U.S.A.

Dr. Jerry Ventre
Former Research Scientist, Florida Solar Energy Center, Oviedo, Florida, U.S.A.

Dr. Richard Wakefield
Vice President, Transmission & Regulatory Services, KEMA, Inc., Fairfax, Virginia, U.S.A.

Contributors

Bill Allemon / *North American Energy Efficiency, Ford Land, Dearborn, Michigan, U.S.A.*

Paul J. Allen / *Reedy Creek Energy Services, Walt Disney World Co., Lake Buena Vista, Florida, U.S.A.*

Fatouh Al-Ragom / *Building and Energy Technologies Department, Kuwait Institute for Scientific Research, Safat, Kuwait*

Kalyan Annamalai / *Department of Mechanical Engineering, College Station, Texas, U.S.A.*

John Archibald / *American Solar Inc., Annandale, Virginia, U.S.A.*

Senthil Arumugam / *Enerquip, Inc., Medford, Wisconsin, U.S.A.*

M. Asif / *School of the Built and Natural Environment, Glasgow Caledonian University, Glasgow, Scotland, U.K.*

Essam Omar Assem / *Arab Fund for Economic and Social Development, Arab Organizations Headquarters Building, Shuwaikh, Kuwait*

U. Atikol / *Energy Research Center, Eastern Mediterranean University, Magusa, Northern Cyprus*

Rick Avery / *Bay Controls, LLC, Maumee, Ohio, U.S.A.*

Lu Aye / *International Technologies Centre (IDTC), Department of Civil and Environmental Engineering, The University of Melbourne, Victoria, Australia*

M. Babcock / *EMO Energy Solutions, LLC, Vienna, Virginia, U.S.A.*

Christopher G. J. Baker / *Chemical Engineering Department, Kuwait University, Safat, Kuwait*

K. Baker / *EMO Energy Solutions, LLC, Vienna, Virginia, U.S.A.*

Peter Barhydt / *Arctic Combustion Ltd., Mississauga, Ontario, Canada*

Larry B. Barrett / *Barrett Consulting Associates, Colorado Springs, Colorado, U.S.A.*

Rosemarie Bartlett / *Pacific Northwest National Laboratory, Richland, Washington, U.S.A.*

Fred Bauman / *Center for the Built Environment, University of California, Berkeley, California, U.S.A.*

David Beal / *Florida Solar Energy Center, Cocoa, Florida, U.S.A.*

Sanford V. Berg / *Director of Water Studies, Public Utility Research Center, University of Florida, Gainesville, Florida, U.S.A.*

Paolo Bertoldi / *European Commission, Directorate General JRC, Ispra (VA), Italy*

Asfaw Beyene / *Department of Mechanical Engineering, San Diego State University, San Diego, California, U.S.A.*

Ujjwal Bhattacharjee / *Energy Engineering Program, University of Massachusetts—Lowell, Lowell, Massachusetts, U.S.A.*

David Bisbee / *Customer Advanced Technologies Program, Sacramento Municipal Utility District (SMUD), Sacramento, California, U.S.A.*

John O. Blackburn / *Department of Economics, Duke University, Durham, North Carolina, U.S.A.*

Richard F. Bonskowski / *Office of Coal, Nuclear, Electric and Alternate Fuels, Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.*

Robert G. Botelho / *Energy Management International, Seekonk, Massachussetts, U.S.A.*

Mark T. Brown / *Department of Environmental Engineering Sciences, University of Florida, Gainesville, Florida, U.S.A.*

Michael L. Brown / *Energy and Environmental Management Center, Georgia Institute of Technology, Savannah, Georgia, U.S.A.*

Alexander L. Burd / *Advanced Research Technology, Suffield, Connecticut, U.S.A.*

Galina S. Burd / *Advanced Research Technology, Suffield, Connecticut, U.S.A.*

James Call / *James Call Engineering, PLLC, Larchmont, New York, U.S.A.*

Norm Campbell / *Energy Systems Group, Newburgh, Indiana, U.S.A.*

Barney L. Capehart / *Department of Industrial and Systems Engineering, University of Florida College of Engineering, Gainesville, Florida, U.S.A.*

Lynne C. Capehart / *Consultant and Technical Writing Specialist, Gainesville, Florida, U.S.A.*

Cristián Cárdenas-Lailhacar / *Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida, U.S.A.*

Jack Casazza / *American Education Institute, Springfield, Virginia, U.S.A.*

Yunus A. Cengel / *Department of Mechanical Engineering, University of Nevada, Reno, Nevada, U.S.A.*

Guangnan Chen / *Faculty of Engineering and Surveying, University of Southern Queensland, Toowoomba, Queensland, Australia*

David E. Claridge / *Department of Mechanical Engineering, and Associate Director, Energy Systems Laboratory, Texas A & M University, College Station, Texas, U.S.A.*

James A. Clark / *Clark Energy, Inc., Broomall, Pennsylvania, U.S.A.*

Gregory Cmar / *Interval Data Systems, Inc., Watertown, Massachusetts, U.S.A.*

Kevin C. Coates / *Transportation & Energy Policy, Coates Consult, Bethesda, Maryland, U.S.A.*

Bruce K. Colburn / *EPS Capital Corp., Villanova, Pennsylvania, U.S.A.*

Stephany L. Cull / *RetroCom Energy Strategies, Inc., Elk Grove, California, U.S.A.*

Rick Cumbo / *Operations and Maintenance Manager, Armstrong Service, Inc., Orlando, Florida, U.S.A.*

Leonard A. Damiano / *EBTRON, Inc., Loris, South Carolina, U.S.A.*

James W. Dean / *Office of Strategic Analysis and Governmental Affairs, Florida Public Service Commission, Tallahassee, Florida, U.S.A.*

Steve DeBusk / *Energy Solutions Manager, CPFilms Inc., Martinsville, Virginia, U.S.A.*

Harvey E. Diamond / *Energy Management International, Conroe, Texas, U.S.A.*

Craig DiLouie / *Lighting Controls Association, National Electrical Manufacturers Association, Rosslyn, Virginia, U.S.A.*

Ibrahim Dincer / *Faculty of Engineering and Applied Science, University of Ontario Institute of Technology (UOIT), Oshawa, Ontario, Canada*

Michael R. Dorrington / *Cameron's Compression Systems Division, Houston, Texas, U.S.A.*

Steve Doty / *Colorado Springs Utilities, Colorado Springs, Colorado, U.S.A.*

David S. Dougan / *EBTRON, Inc., Loris, South Carolina, U.S.A.*

John Duffy / *Department of Mechanical Engineering, University of Massachussetts—Lowell, Lowell, Massachussetts, U.S.A.*

Philip Fairey / *Florida Solar Energy Center, Cocoa, Florida, U.S.A.*

Carla Fair-Wright / *Maintenance Technology Services, Cameron's Compression Systems Division, Houston, Texas, U.S.A.*

Ahmad Faruqui / *The Brattle Group, San Francisco, California, U.S.A.*

J. Michael Fernandes / *Mobile Platforms Group, Intel Corporation, Santa Clara, California, U.S.A.*

Steven Ferrey / *Suffolk University Law School, Boston, Massachusetts, U.S.A.*

Andrew R. Forrest / *Government Communications Systems Division, Harris Corporation, Melbourne, Florida, U.S.A.*

R. Scot Foss / *IR Air Solutions, Davidson, North Carolina, U.S.A.*

Jackie L. Francke / *Geotechnika, Inc., Longmont, Colorado, U.S.A.*

Fred Freme / *Office of Coal, Nuclear, Electric and Alternate Fuels, Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.*

Dwight K. French / *Energy Consumption Division, Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.*

Kevin Fuller / *Interval Data Systems, Inc., Watertown, Massachusetts, U.S.A.*

Kamiel S. Gabriel / *University of Ontario Institute of Technology, Oshawa, Ontario, Canada*

Clark W. Gellings / *Electric Power Research Institute (EPRI), Palo Alto, California, U.S.A.*

Geoffrey J. Gilg / *Pepco Energy Services, Inc., Arlington, Virginia, U.S.A.*

Bill Gnerre / *Interval Data Systems, Inc., Watertown, Massachusetts, U.S.A.*

Fredric S. Goldner / *Energy Management and Research Associates, East Meadow, New York, U.S.A.*

Dan Golomb / *Department of Environmental, Earth and Atmospheric Sciences, University of Massachusetts—Lowell, Lowell, Massachusetts, U.S.A.*

John Van Gorp / *Power Monitoring and Control, Schneider Electric, Saanichton, British Columbia, Canada*

Alex E. S. Green / *Green Liquids and Gas Technologies, Gainesville, Florida, U.S.A.*

David C. Green / *Green Management Services, Inc., Fort Myers, Florida, U.S.A.*

L. J. Grobler / *School for Mechanical Engineering, North-West University, Potchefstroom, South Africa*

H. M. Güven / *Energy Research Center, Eastern Mediterranean University, Magusa, Northern Cyprus*

Mark A. Halverson / *Pacific Northwest National Laboratory, West Halifax, Vermont, U.S.A.*

Shirley J. Hansen / *Hansen Associates, Inc., Gig Harbor, Washington, U.S.A.*

Susanna S. Hanson / *Global Applied Systems, Trane Commercial Systems, La Crosse, Wisconsin, U.S.A.*

Glenn C. Haynes / *RLW Analytics, Middletown, Connecticut, U.S.A.*

Warren M. Heffington / *Industrial Assessment Center, Department of Mechanical Engineering, Texas A&M University, College Station, Texas, U.S.A.*

Arif Hepbasli / *Department of Mechanical Engineering, Faculty of Engineering, Ege University, Izmir, Turkey*

Keith E. Herold / *Department of Mechanical Engineering, University of Maryland, College Park, Maryland, U.S.A.*

James G. Hewlett / *Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.*

Richard F. Hirsh / *Department of History, Virginia Tech, Blacksburg, Virginia, U.S.A.*

Kevin Hoag / *Engine Research Center, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.*

Nathan E. Hultman / *Science, Technology, and International Affairs, Georgetown University, Washington, D.C., U.S.A.*

W. David Hunt / *Pacific Northwest National Laboratory, Washington, D.C., U.S.A.*

H. A. Ingley III / *Department of Mechanical Engineering, University of Florida, Gainesville, Florida, U.S.A.*

Nevena H. Iordanova / *Senior Utility Systems Engineer, Armstrong Service, Inc., Orlando, Florida, U.S.A.*

Charles P. Ivey / *Aiken Global Environmental/FT Benning DPW, Fort Benning, Georgia, U.S.A.*

Mark A. Jamison / *Public Utility Research Center, University of Florida, Gainesville, Florida, U.S.A.*

Somchai Jiajitsawat / *Department of Mechanical Engineering, University of Massachussetts—Lowell, Lowell, Massachussetts, U.S.A.*

Gholamreza Karimi / *Department of Mechanical Engineering, University of Waterloo, Waterloo, Ontario, Canada*

Janey Kaster / *Yamas Controls, Inc., South San Francisco, California, U.S.A.*

Sila Kiliccote / *Commercial Building Systems Group, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.*

Birol I. Kilkis / *Green Energy Systems, International LLC, Vienna, Virginia, U.S.A.*

Ronald L. Klaus / *VAST Power Systems, Elkhart, Indiana, U.S.A.*

Milivoje M. Kostic / *Department of Mechanical Engineering, Northern Illinois University, DeKalb, Illinois, U.S.A.*

Athula Kulatunga / *Department of Electrical and Computer Engineering Technology, Purdue University, West Lafayette, Indiana, U.S.A.*

James W. Leach / *Industrial Assessment Center, Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, North Carolina, U.S.A.*

Larry Leetzow / *World Institute of Lighting and Development Corp./Magnaray International Division, Sarasota, Florida, U.S.A.*

Jim Lewis / *Obvius LLC, Hillsboro, Oregon, U.S.A.*

Xianguo Li / *Department of Mechanical Engineering, University of Waterloo, Waterloo, Ontario, Canada*

Todd A. Litman / *Victoria Transport Policy Institute, Victoria, British Columbia, Canada*

Mingsheng Liu / *Architectural Engineering Program, Director, Energy Systems Laboratory, Peter Kiewit Institute, University of Nebraska—Lincoln, Omaha, Nebraska, U.S.A.*

Robert Bruce Lung / *Resource Dynamics Corporation, Vienna, Virginia, U.S.A.*

Alfred J. Lutz / *AJL Resources, LLC, Philadelphia, Pennsylvania, U.S.A.*

David MacPhaul / *CH2M HILL, Gainesville, Florida, U.S.A.*

Richard J. Marceau / *University of Ontario Institute of Technology, Oshawa, Ontario, Canada*

Rudolf Marloth / *Department of Mechanical Engineering, Loyola Marymount University, Los Angeles, California, U.S.A.*

Sandra B. McCardell / *Current-C Energy Systems, Inc., Mills, Wyoming, U.S.A.*

William Ross McCluney / *Florida Solar Energy Center, University of Central Florida, Cocoa, Florida, U.S.A.*

Donald T. McGillis / *Pointe Claire, Québec, Canada*

Aimee McKane / *Lawrence Berkeley National Laboratory, Washington, D.C., U.S.A.*

D. Paul Mehta / *Department of Mechanical Engineering, Bradley University, Peoria, Illinois, U.S.A.*

Zohrab Melikyan / *HVAC Department, Armenia State University of Architecture and Construction, Yerevan, Armenia*

Ingrid Melody / *Florida Solar Energy Center, Cocoa, Florida, U.S.A.*

Mark Menezes / *Emerson Process Management (Rosemount), Mississauga, Ontario, Canada*

Adnan Midilli / *Department of Mechanical Engineering, Faculty of Engineering, Nigde University, Nigde, Turkey*

Ruth Mossad / *Faculty of Engineering and Surveying, University of Southern Queensland, Toowoomba, Queensland, Australia*

Naoya Motegi / *Commercial Building Systems Group, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.*

Alex V. Moultanovsky / *Automotive Climate Control, Inc., Elkhart, Indiana, U.S.A.*

Martin A. Mozzo / *M and A Associates, Inc., Robbinsville, New Jersey, U.S.A.*

T. Muneer / *School of Engineering, Napier University, Edinburgh, Scotland, U.K.*

Greg F. Naterer / *University of Ontario Institute of Technology, Oshawa, Ontario, Canada*

Robert R. Nordhaus / *Van Ness Feldman, PC, Washington, D.C., U.S.A.*

Michael M. Ohadi / *The Petroleum Institute, Abu Dhabi, United Arab Emirates*

K. E. Ohrn / *Cypress Digital Ltd., Vancouver, British Columbia, Canada*

Svetlana J. Olbina / *Rinker School of Building Construction, University of Florida, Gainesville, Florida, U.S.A.*

Eric Oliver / *EMO Energy Solutions, LLC, Vienna, Virginia, U.S.A.*

Mitchell Olszewski / *Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.*

Bohdan W. Oppenheim / *U.S. Department of Energy Industrial Assessment Center, Loyola Marymount University, Los Angeles, U.S.A.*

Leyla Ozgener / *Department of Mechanical Engineering, Celal Bayar University, Manisa, Turkey*

Graham Parker / *Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.*

Steven A. Parker / *Pacific Northwest National Laboratory, Richland, Washington, U.S.A.*

Ken E. Patterson II / *Advanced Energy Innovations—Bringing Innovation to Industry, Menifee, California, U.S.A.*

Klaus-Dieter E. Pawlik / *Accenture, St Petersburg, Florida, U.S.A.*

David E. Perkins / *Active Power, Inc., Austin, Texas, U.S.A.*

Jeffery P. Perl / *Chicago Chem Consultants Corp., Chicago, Illinois, U.S.A.*

Neil Petchers / *NORESCO Energy Consulting, Shelton, Connecticut, U.S.A.*

Eric Peterschmidt / *Honeywell Building Solutions, Honeywell, Golden Valley, Minnesota, U.S.A.*

Robert W. Peters / *Department of Civil, Construction & Environmental Engineering, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.*

Mark A. Peterson / *Sustainable Success LLC, Clementon, New Jersey, U.S.A.*

Mary Ann Piette / *Commercial Building Systems Group, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.*

Wendell A. Porter / *University of Florida, Gainesville, Florida, U.S.A.*

Soyuz Priyadarsan / *Texas A&M University, College Station, Texas, U.S.A.*

Sam Prudhomme / *Bay Controls, LLC, Maumee, Ohio, U.S.A.*

Jianwei Qi / *Department of Mechanical Engineering, University of Maryland, College Park, Maryland, U.S.A.*

Ashok D. Rao / *Advanced Power and Energy Program, University of California, Irvine, California, U.S.A.*

Ab Ream / *U.S. Department of Energy, Washington, D.C., U.S.A.*

Rolf D. Reitz / *Engine Research Center, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.*

Rich Remke / *Commercial Systems and Services, Carrier Corporation, Syracuse, New York, U.S.A.*

Silvia Rezessy / *Energy Efficiency Advisory, REEEP International Secretariat, Vienna International Centre, Austria, Environmental Sciences and Policy Department, Central European University, Nador, Hungary*

James P. Riordan / *Energy Department, DMJM + Harris / AECOM, New York, U.S.A.*

Vernon P. Roan / *University of Florida, Gainesville, Florida, U.S.A.*

Stephen A. Roosa / *Energy Systems Group, Inc., Louisville, Kentucky, U.S.A.*

Michael Ropp / *South Dakota State University, Brookings, South Dakota, U.S.A.*

Marc A. Rosen / *Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, Ontario, Canada*

Christopher Russell / *Energy Pathfinder Management Consulting, LLC, Baltimore, Maryland, U.S.A.*

Hemmat Safwat / *Project Development Power and Desalination Plants, Consolidated Contractors International Company, Athens, Greece*

Abdou-R. Sana / *Montreal, Québec, Canada*

Diane Schaub / *Industrial and Systems Engineering, University of Florida, Gainesville, Florida, U.S.A.*

Diana L. Shankle / *Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.*

S. A. Sherif / *Department of Mechanical and Aerospace Engineering, Wayne K. and Lyla L. Masur HVAC Laboratory, University of Florida, Gainesville, Florida, U.S.A.*

Brian Silvetti / *CALMAC Manufacturing Corporation, Fair Lawn, New Jersey, U.S.A.*

Carey J. Simonson / *Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*

Amy Solana / *Pacific Northwest National Laboratory, Portland, Oregon, U.S.A.*

Paul M. Sotkiewicz / *Public Utility Research Center, University of Florida, Warrington College of Business, Gainesville, Florida, U.S.A.*

Amanda Staudt / *Board on Atmospheric Sciences and Climate, The National Academies, Washington, D.C., U.S.A.*

Nick Stecky / *NJS Associates, Denville, New Jersey, U.S.A.*

Kate McMordie Stoughton / *Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.*

Therese Stovall / *Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.*

Gregory P. Sullivan / *Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.*

John M. Sweeten / *Texas Agricultural Experiment Station, Amarillo, Texas, U.S.A.*

Thomas F. Taranto / *ConservAIR Technologies, LLP, Baldwinsville, New York, U.S.A.*

Michael Taylor / *Honeywell Building Solutions, Honeywell, St. Louis Park, Minnesota, U.S.A.*

Albert Thumann / *Association of Energy Engineers, Atlanta, Georgia, U.S.A.*

Robert J. Tidona / *RMT, Incorporated, Plymouth Meeting, Pennsylvania, U.S.A.*

Jill S. Tietjen / *Technically Speaking, Inc., Greenwood Village, Colorado, U.S.A.*

Lorrie B. Tietze / *Interface Consulting, LLC, Castle Rock, Colorado, U.S.A.*

Greg Tinkler / *RLB Consulting Engineers, Houston, Texas, U.S.A.*

Alberto Traverso / *Dipartimento di Macchine Sistemi Energetici e Trasporti, Thermochemical Power Group, Università di Genova, Genova, Italy*

Douglas E. Tripp / *Canadian Institute for Energy Training, Rockwood, Ontario, Canada*

James S. Tulenko / *Laboratory for Development of Advanced Nuclear Fuels and Materials, University of Florida, Gainesville, Florida, U.S.A.*

W. D. Turner / *Director, Energy Systems Laboratory, Texas A & M University, College Station, Texas, U.S.A.*

Wayne C. Turner / *Industrial Engineering and Management, Oklahoma State University, Stillwater, Oklahoma, U.S.A.*

S. Kay Tuttle / *Process Heating, Boilers and Foodservices, Duke Energy, Charlotte, North Carolina, U.S.A.*

Robert E. Uhrig / *University of Tennessee, Knoxville, Tennessee, U.S.A.*

Sergio Ulgiati / *Department of Sciences for the Environment, Parthenope, University of Naples, Napoli, Italy*

Francisco L. Valentine / *Premier Energy Services, LLC, Columbia, Maryland, U.S.A.*

Ing. Jesús Mario Vignolo / *Instituto Ingeniería Eléctrica, Universidad de la República, Montevideo, Uruguay*

Paul L. Villeneuve / *University of Maine, Orono, Maine, U.S.*

T. G. Vorster / *School for Mechanical Engineering, North-West University, Potchefstroom, South Africa*

James P. Waltz / *Energy Resource Associates, Inc., Livermore, California, U.S.A.*

Devra Bachrach Wang / *Natural Resources Defense Council, San Francisco, California, U.S.A.*

John W. Wang / *Department of Mechanical Engineering, University of Massachussetts—Lowell, Lowell, Massachussetts, U.S.A.*

David S. Watson / *Commercial Building Systems Group, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.*

William D. Watson / *Office of Coal, Nuclear, Electric and Alternate Fuels, Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.*

Marty Watts / *V-Kool, Inc., Houston, Texas, U.S.A.*

Tom Webster / *Center for the Built Environment, University of California, Berkeley, California, U.S.A.*

Michael K. West / *Building Systems Scientists, Advantek Consulting, Inc., Melbourne, Florida, U.S.A.*

Robert E. Wilson / *ConservAIR Technologies Co., LLP, Kenosha, Wisconsin, U.S.A.*

Michael Burke Wood / *Ministry of Energy, Al Dasmah, Kuwait*

Kurt E. Yeager / *Galvin Electricity Initiative, Palo Alto, California, U.S.A.*

Contents

Volume 1

Volume 2

Volume 3

Topical Table of Contents

I. Energy, Energy Sources, and Energy Use

I. Energy, Energy Sources, and Energy Use (*cont'd.*)

Derived Energy (cont'd.)

Energy Storage

Fuel Cells

II. Principles of Energy Analysis and Systems/Economic Analysis of Energy Systems

Energy Systems Analysis

Financial and Economic Analysis Principles

Principles of Electric Energy Systems

Principles of Thermal Energy Systems

III. Utilities, Suppliers of Energy, and Utility Regulation

Electric Supply System

Fuel Supply System

Utilities: Overview

Utility Regulatory Issues

IV. Facilities and Users of Energy

Building Envelope

IV. Facilities and Users of Energy (*cont'd.*)

Building Envelope (cont'd.)

Compressed Air Systems

Facility Controls and Information Systems

Heat Recovery

HVAC Systems

Industrial Facilities

Lighting and Motors

On Site Electric Generation

Space Heating, Water Heating, and Water Use

Steam Boilers and Steam Systems

V. Energy Management

Commissioning

Energy Auditing and Benchmarking

Energy Codes and Standards

Energy Management Programs for Facilities

Energy Savings Projects

Measurement and Verification of Energy Savings

Performance Contracting

VI. Current Energy and Environmental Issues

Environmental Regulation: Overview

Sustainable Buildings

VII. Transportation and Other Energy Uses

Other Energy Uses

VII. Transportation and Other Energy Uses (*cont'd.*)

Transportation

Foreword

The Association of Energy Engineers (AEE) is proud to be a sponsor of the *Encyclopedia of Energy Engineering and Technology, Three-Volume Set*, edited by Dr. Barney L. Capehart. In 2007 AEE is celebrating its 30th anniversary and it is a fitting tribute that the *Encyclopedia of Energy Engineering and Technology* is published at this time. AEE defined the energy engineering profession and this comprehensive work details the core elements for success in this field.

Dr. Capehart has performed a monumental task of facilitating over 300 researchers and practitioners who have contributed to this three-volume set. These distinguished authorities share a wealth of knowledge on approximately 190 topics. Dr. Capehart, through his training and publications, has significantly impacted the energy engineering profession and has helped make it what it is today.

Global climate change concerns and unstable energy prices have raised the importance of energy engineering. This encyclopedia will be a valuable tool in assisting energy engineers to reach their potential. The Association of Energy Engineers (AEE) and our network of 8,000 members in 77 countries would like to thank Dr. Barney Capehart and the numerous volunteers who have made this work possible.

Albert Thumann, P.E., CEM
Executive Director
The Association of Energy Engineers (AEE)

Preface

Energy engineers and technologists have made efficient and cost effective devices for many years which provide the energy services society wants and expects. From air conditioners to waste fuels, energy engineers and technologists continue to make our lives comfortable and affordable using limited resources in efficient and renewable ways.

Over 300 researchers and practitioners, through 190 entries, provide ready access to the basic principles and applications of energy engineering, as well as advanced applications in the technologies of energy production and use. The global supply of energy is increasingly being stressed to provide for an expanding world population. Energy efficiency, energy conservation through energy management, and use of renewable energy sources are three of the major strategies that in the future will help provide the energy and energy services for the world's population and the world's economy.

This unique reference contains state-of-the-art progress on the most important topics in this field of energy engineering and technology. All entries in the encyclopedia have been written by experts in their specialties, and have been reviewed by subject matter authorities. This distinguished group of experts share a wealth of knowledge on topics such as:

- Energy, energy supplies and energy use
- Renewable and alternative energy sources
- Technical, economic and financial analysis of energy systems
- Energy uses in buildings and industry
- Energy efficiency and energy conservation opportunities and projects
- Commissioning, benchmarking, performance contracting, and measurement and verification
- Environmental regulation and public policy for energy supply and use
- Global climate change and carbon control
- Sustainable buildings and green development
- Hybrid electric and hydrogen fueled vehicles and maglev transportation

The *Encyclopedia of Energy Engineering and Technology, Three-Volume Set,* is a key reference work for professionals in academia, business, industry and government, as well as students at all levels. It should be regularly consulted for basic and advanced information to guide students, scholars, practitioners, the public, and policy makers. Contributions address a wide spectrum of theoretical and applied topics, concepts, methodologies, strategies, and possible solutions.

The On-Line Edition is a dynamic resource that will grow as time and knowledge progress. Suggestions for additional content are welcomed by the editor, and new authors should contact me at the e-mail address listed below.

As editor, I would like to thank the people who worked hard to initiate this encyclopedia, and make the project a success. Thanks go to Russell Dekker at Marcel Dekker, and Al Thumann, the Executive Director of the Association of Energy Engineers (AEE) for getting the project going. I appreciate their confidence in my ability to accomplish this immense project. Part of the purpose of this project is to help provide professional and educational support for new people coming into our area of energy engineering. A profession can only succeed and grow if new people have a resource to learn about a new area, and find this area interesting and exciting for their careers.

Directors Oona Schmidt and Claire Miller were both excellent in helping to organize and specify the work that had to be done to get the encyclopedia started and on track for completion. Editorial Assistants Andrea Cunningham, Lousia Lam, and Marisa Hoheb provided daily help to me, and kept all of the records and contacts with the authors. Their help was invaluable.

Preparation of this modern compendium on energy engineering and technology has only been possible through the commitment and hard work of hundreds of energy engineers from around the globe. I want to thank all of the

authors for their outstanding efforts to identify major topics of interest for this project, and to write interesting and educational articles based on their areas of expertise. Many of the authors also served a dual function of both writing their own articles, and reviewing the submissions of other authors. Another important group of people were those on the Editorial Board who helped submit topics, organizational ideas, and lists of potential authors for the encyclopedia. This Board was a great help in getting the actual writing of articles started, as well as many of the Editorial Board members also contributed articles themselves.

I would also like to thank my wife Lynne for her continuing support of all of my projects over the years. And finally, I would like to dedicate this encyclopedia to my grandchildren Hannah and Easton. They are a part of the future that I hope will be using the efficient and sustainable energy resources presented in this encyclopedia.

Barney L. Capehart, Editor
Capehart@ise.ufl.edu

Common Energy Abbreviations

The following abbreviations were provided by the Energy Information Administration's National Energy Information Center. The Energy Information Administration is a statistical agency of the U.S. Department of Energy.

AC	alternating current
AFUDC	allowance for funds used during construction
AFV	alternative-fuel vehicle
AGA	American Gas Association
ANSI	American National Standards Institute
API	American Petroleum Institute
ASTM	American Society for Testing and Materials
bbl	barrel(s)
bbl/d	barrel(s) per day
bcf	billion cubic feet
BLS	Bureau of Labor Statistics within the U.S. Department of Labor
BOE	barrels of oil equivalent (used internationally)
Btu	British thermal unit(s)
BWR	boiling-water reactor
C/gal	cents per gallon
CAFE	corporate average fuel economy
CARB	California Air Resources Board
CDD	cooling degree-days
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CF	cubic foot
CFC	chlorofluorocarbon
CFS	cubic feet per second
CH_4	Methane
CHP	combined heat and power
CNG	compressed natural gas
cnt	cent
CO	carbon monoxide
CO_2	carbon dioxide
CPI	consumer price index
CWIP	construction work in progress
DC	direct current
DOE	Department of Energy
DRB	demonstrated reserve base
DSM	demand-side management
E85	A fuel containing a mixture of 85 percent ethanol and 15 percent gasoline
E95	A fuel containing a mixture of 95 percent ethanol and 5 percent gasoline
EAR	estimated additional resources
EIA	Energy Information Administration
EIS	Environmental Impact Statement
EOR	enhanced oil recovery
EPA	Environmental Protection Agency
EPACT	Energy Policy Act
EU	European Union
EWG	exempt wholesale generator
FASB	Financial Accounting Standards Board
FBR	fast breeder reactor

FERC	Federal Energy Regulatory Commission
FGD	flue-gas desulfurization
F.O.B	free on board
FPC	Federal Power Commission
FRS	Financial Reporting System
gal	gallon
GDP	gross domestic product
GNP	gross national product
GVW	gross vehicle weight
GW	gigawatt
GWe	gigawatt-electric
GWh	gigawatthour
GWP	global warming potential
HCFC	hydrochlorofluorocarbon
HDD	heating degree-days
HFC	hydrofluorocarbon
HID	high-intensity discharge
HTGR	high temperature gas-cooled reactor
HVAC	heating, ventilation, and air-conditioning
IEA	International Energy Agency
IOU	investor-owned utility
IPP	independent power producer
ISO	independent system operator
kVa	kilovolt-Ampere
kW	kilowatt
kWe	kilowatt-electric
kWh	kilowatthour
lb	pound
LDC	local distribution company
LEVP	Low Emissions Vehicle Program
LHV	lower heating value
LIHEAP	Low-Income Home Energy Assistance Program
LNG	liquefied natural gas
LPG	liquefied petroleum gases
LRG	liquefied refinery gases
LWR	light water reactor
M	thousand
Mcf	one thousand cubic feet
MECS	Manufacturing Energy Consumption Survey
MM	million (10^6)
MMbbl/d	one million (10^6) barrels of oil per day
MMBtu	one million (10^6) British thermal units
MMcf	one million (10^6) cubic feet
MMgal/d	one million (10^6) gallons per day
MMst	one million (10^6) short tons
MPG	miles per gallon
MSA	metropolitan statistical area
MSHA	Mine Safety and Health Administration
MSW	municipal solid waste
MTBE	methyl tertiary butyl ether
MW	megawatt
MWe	megawatt electric
MWh	megawatthour
N_2O	Nitrous oxide
NAAQS	National Ambient Air Quality Standards
NAICS	North American Industry Classification System

NARUC	National Association of Regulatory Utility Commissioners
NERC	North American Electric Reliability Council
NGL	natural gas liquids
NGPA	Natural Gas Policy Act of 1978
NGPL	natural gas plant liquids
NGV	natural gas vehicle
NOAA	National Oceanic and Atmospheric Administration
NOPR	Notice of Proposed Rulemaking
NO_x	nitrogen oxides
NRECA	National Rural Electric Cooperative Association
NUG	nonutility generator
NURE	National Uranium Resource Evaluation
NYMEX	New York Mercantile Exchange
O_3	Ozone
O&M	operation and maintenance
OCS	Outer Continental Shelf
OECD	Organization for Economic Cooperation and Development
OEM	original equipment manufacturers
OPEC	Organization of Petroleum Exporting Countries
OPRG	oxygenated fuels program reformulated gasoline
OTEC	ocean thermal energy conversion
PADD	Petroleum Administration for Defense Districts
PBR	pebble-bed reactor
PBR	performance-based rates
PCB	polychlorinated biphenyl
PFCs	perfluorocarbons
PGA	purchased gas adjustment
PPI	Producer Price Index
PUD	public utility district
PUHCA	Public Utility Holding Company Act of 1935
PURPA	Public Utility Regulatory Policies Act of 1978
PV	photovoltaic
PVC	photovoltaic cell; polyvinyl chloride
PWR	pressurized-water reactor
QF	qualifying facility
QUAD	quadrillion Btu: 10^{15} Btu
RAC	refiners' acquisition cost
RAR	reasonable assured resources
RBOB	Reformulated Gasoline Blendstock for Oxygenate Blending
RDF	refuse-derived fuel
REA	Rural Electrification Administration
RECS	Residential Energy Consumption Survey
RFG	reformulated gasoline
RSE	relative standard error
RVP	Reid vapor pressure
SEER	seasonal energy efficiency ratio
SF_6	sulfur hexafluoride
SI	International System of Units (Système international d'unités)
SIC	Standard Industrial Classification
SNG	synthetic natural gas
SO_2	sulfur dioxide
SPP	small power producer
SPR	Strategic Petroleum Reserve
SR	speculative resources
T	trillion 10^{12}
TVA	Tennessee Valley Authority

TW	Terawatt
U_3O_8	Uranium oxide
UF_6	uranium hexaflouride
ULCC	ultra large crude carrier
UMTRA	Uranium Mill Tailings Radiation Control Act of 1978
USACE	U.S. Army Corps of Engineers (sometimes shortened to USCE in EIA tables)
USBR	United States Bureau of Reclamation
V	Volt
VAWT	vertical-axis wind turbine
VIN	vehicle identification number
VLCC	very large crude carrier
VMT	vehicle miles traveled
VOC	volatile organic compound
W	watt
WACOG	weighted average cost of gas
Wh	watt hour
WTI	West Texas Intermediate

Thermal Metric and Other Conversion Factors

The following tables appeared in the January 2007 issue of the Energy Information Administration's *Monthly Energy Review*. The Energy Information Administration is a statistical agency of the U.S. Department of Energy.

Table 1 Metric Conversion Factors
These metric conversion factors can be used to calculate the metric-unit equivalents of values expressed in U.S. Customary units. For example, 500 short tons are the equivalent of 453.6 metric tons (500 short tons × 0.9071847 metric tons/short ton = 453.6 metric tons).

Type of Unit	U.S. Unit		Equivalent in Metric Units	
Mass	1 short ton (2,000 lb)	=	0.907 184 7	metric tons (t)
	1 long ton	=	1.016 047	metric tons (t)
	1 pound (lb)	=	0.453 592 37[a]	kilograms (kg)
	1 pound uranium oxide (lb U^3O^8)	=	0.384 647[b]	kilograms uranium (kgU)
	1 ounce, avoirdupois (avdp oz)	=	28.349 52	grams (g)
Volume	1 barrel of oil (bbl)	=	0.158 987 3	cubic meters (m^3)
	1 cubic yard (yd^3)	=	0.764 555	cubic meters (m^3)
	1 cubic foot (ft^3)	=	0.028 316 85	cubic meters (m^3)
	1 U.S. gallon (gal)	=	3.785 412	liters (L)
	1 ounce, fluid (fl oz)	=	29.573 53	milliliters (mL)
	1 cubic inch (in^3)	=	16.387 06	milliliters (mL)
Length	1 mile (mi)	=	1.609 344[a]	kilometers (km)
	1 yard (yd)		0.914 4[a]	meters (m)
	1 foot (ft)	=	0.304 8[a]	meters (m)
	1 inch (in)	=	2.54[a]	centimeters (cm)
Area	1 acre	=	0.404 69	hectares (ha)
	1 square mile (mi^2)	=	2.589 988	square kilometers (km^2)
	1 square yard (yd^2)	=	0.836 127 4	square meters (m^2)
	1 square foot (ft^2)	=	0.092 903 04[a]	square meters (m^2)
	1 square inch (in^2)	=	6.451 6[a]	square centimeters (cm^2)
Energy	1 British thermal unit (Btu)[c]	=	1,055.055 852 62[a]	joules (J)
	1 calorie (cal)	=	4.186 8[a]	joules (J)
	1 kilowatthour (kWh)	=	3.6[a]	megajoules (MJ)
Temperature[d]	32 degrees Fahrenheit (°F)	=	0[a]	degrees Celsius (°C)
	212 degrees Fahrenheit (°F)	=	100[a]	degrees Celsius (°C)

[a]Exact conversion.
[b]Calculated by the Energy Information Administration.
[c]The Btu used in this table is the International Table Btu adopted by the Fifth International Conference on Properties of Steam, London, 1956.
[d]To convert degrees Fahrenheit (°F) to degrees Celsius (°C) exactly, subtract 32, then multiply by 5/9.

Notes: • Spaces have been inserted after every third digit to the right of the decimal for ease of reading. • Most metric units belong to the International System of Units (SI), and the liter, hectare, and metric ton are accepted for use with the SI units. For more information about the SI units, see http://physics.nist.gov/cuu/Units/index.html.
Web Page: http://www.eia.doe.gov/emeu/mer/append_b.html.
Sources: • General Services Administration, Federal Standard 376B, *Preferred Metric Units for General Use by the Federal Government* (Washington, D.C., January 1993), pp. 9-11, 13, and 16. • U.S. Department of Commerce, National Institute of Standards and Technology, Special Publications 330, 811, and 814. • American National Standards Institute/ Institute of Electrical and Electronic Engineers, ANSI/IEEE Std 268-1992, pp. 28 and 29.

Table 2 Metric Prefixes
The names of multiples and subdivisions of any unit may be derived by combining the name of the unit with prefixes as below.

Unit Multiple	Prefix	Symbol	Unit Subdivision	Prefix	Symbol
10^1	deka	da	10^{-1}	deci	D
10^2	hecto	h	10^{-2}	centi	c
10^3	kilo	k	10^{-3}	milli	m
10^6	mega	M	10^{-6}	micro	μ
10^9	giga	G	10^{-9}	nano	n
10^{12}	tera	T	10^{-12}	pico	p
10^{15}	peta	P	10^{-15}	femto	f
10^{18}	exa	E	10^{-18}	atto	a
10^{21}	zetta	Z	10^{-21}	zepto	z
10^{24}	yotta	Y	10^{-24}	yocto	y

Web Page: http://www.eia.doe.gov/emeu/mer/append_b.html.
Source: U.S. Department of Commerce, National Institute of Standards and Technology, *The International System of Units (SI)*, NIST Special Publication 330, 1991 Edition (Washington, D.C., August 1991), p. 10.

Table 3 Other Physical Conversion Factors
The factors below can be used to calculate equivalents in various physical units commonly used in energy analyses. For example, 10 barrels are the equivalent of 420 U.S. gallons (10 barrels x 42 gallons/barrel = 420 gallons).

Energy Source	Original Unit		Equivalent in Final Units	
Petroleum	1 barrel (bbl)	=	42[a]	U.S. gallons (gal)
Coal	1 short ton	=	2,000[a]	pounds (lb)
	1 long ton	=	2,240[a]	pounds (lb)
	1 metric ton (t)	=	1,000[a]	kilograms (kg)
Wood	1 cord (cd)	=	1.25[b]	shorts tons
	1 cord (cd)	=	128[a]	cubic feet (ft^3)

[a]Exact conversion.
[b]Calculated by the Energy Information Administration.
Web Page: http://www.eia.doe.gov/emeu/mer/append_b.html.
Source: U.S. Department of Commerce, National Institute of Standards and Technology, *Specifications, Tolerances, and Other Technical Requirements for Weighing and Measuring Devices*, NIST Handbook 44, 1994 Edition (Washington, D.C., October 1993), pp. B-10, C-17 and C-21.

About the Editor

Dr. Barney L. Capehart is a Professor Emeritus of Industrial and Systems Engineering at the University of Florida, Gainesville, Florida. He has BS and Master's degrees in Electrical Engineering and a Ph.D. in Systems Engineering. He taught at the University of Florida from 1968 until 2000, where he taught a wide variety of courses in energy systems analysis, energy efficiency and computer simulation. For the last 25–30 years, energy systems analysis has been his main area of research and publication. He is the co-author or chapter contributor on six books on energy topics, the author of over 50 energy research articles in scholarly journals, and has been involved in 11–12 funded energy research projects totaling over $2,000,000.

He has performed energy efficiency and utility research projects for the Florida Governor's Energy Office, the Florida Public Service Commission, and several utilities in the state of Florida. In addition, he has served as an Expert Witness in the development of the Florida Energy Efficiency and Conservation Act electric growth goals, the Florida Rules for Payments to Cogenerators, and the passage of the Florida Appliance Efficiency Standards Act. He is one of the state's leading experts on electric utility Demand Side Management programs for reducing customer costs and for increasing the efficiency of customer end-use.

He also has broad experience in the commercial/industrial sector, with his management of the University of Florida Industrial Assessment Center from 1990 until 1999. This center performs audits for small and medium sized manufacturing plants, and is funded by the U.S. Department of Energy. He has personally conducted over 100 audits of industrial facilities, and has helped students conduct audits of hundreds of office buildings, small businesses, government facilities, and apartment complexes. He regularly taught a University of Florida course on Energy Management to about fifty engineering students each year. He currently teaches Energy Management seminars around the country and around the world for the Association of Energy Engineers (AEE). In addition, he has been a prolific author of newspaper articles and has appeared on numerous radio and TV talk shows to discuss his work on energy efficiency and reducing people's electric bills. He is also a popular speaker at civic, professional, educational, and government meetings to talk about the benefits of energy efficiency.

Dr. Capehart is a Fellow of the Institute of Electrical and Electronic Engineers, the American Association for the Advancement of Science, and the Institute of Industrial Engineers. He is also a Member of the Hall of Fame of the AEE. He is listed in Who's Who in the World and Who's Who in America. In 1988 he was awarded the Palladium Medal by the American Association of Engineering Societies for his work on energy systems analysis and appliance efficiency standards. He has conducted many energy research projects, and has published over fifty journal articles. He was the Editor of the *International Journal of Energy Systems* from 1985–1988.

He is the editor of *Information Technology for Energy Managers—Understanding Web Based Energy Information and Control Systems*, Fairmont Press, 2004; senior co-editor of *Web Based Energy Information and Control Systems*, Fairmont Press, 2005; senior co-editor of *Web Based Enterprise Energy Management and BAS Systems*, Fairmont Press, 2007; senior co-author of the fifth edition of the *Guide to Energy Management*, Fairmont Press, 2006; and author of the chapter on *Energy Management* in the *Handbook of Industrial Engineering, Second Edition*, by Gavriel Salvendy. He also wrote the chapter on *Energy Auditing* for the *Energy Management Handbook, Sixth Edition* by Wayne C. Turner.

Encyclopedia of

Energy Engineering and Technology

First Edition

Volume 1

Pages 1 through 572

Acc–Energy Serv

Accounting: Facility Energy Use

Douglas E. Tripp
Canadian Institute for Energy Training, Rockwood, Ontario, Canada

Abstract
Energy Accounting is the management technique that quantitatively monitors energy consumption, relates consumption to key independent variables such as production and weather, and assesses energy performance or efficiency over time and against relevant benchmarks. The successful practice of energy accounting is predicated on the identification of the right kinds of data to be collected, the use of appropriate statistical methods to correlate consumption to the independent variables, and the reporting of the right information to the right people in the organization. Energy Monitoring and Targeting (M&T) is a technique for energy performance analysis that overcomes possible deficiencies in the traditional performance indices or energy intensity.

INTRODUCTION

Energy Accounting is an essential component of effective energy management, just as financial accounting is essential to organizational management. In order to gain the full benefit of energy management, organizations need to be able to monitor their energy consumption, relate consumption to the independent variables that drive it, compare the energy performance of their plants and buildings to themselves over time and to other similar facilities, and assess the impact of energy saving measures.

This description of energy accounting is intended to (1) provide insight into the basic principles and methods of energy accounting and (2) expand the conventional understanding of energy accounting to include the analysis technique commonly known as Monitoring and Targeting.

Monitoring, recording, and reporting on gross energy consumption are all straightforward. Complications arise, however, when performance indices enter the picture. Indices in the industrial environment referred to collectively as energy intensity relate energy consumption to measures of production, for example GJ/tonne, MBtu/ton, and so on. Performance indices that are relevant to building operations typically relate energy consumption to conditioned floor area. For reasons developed in this article, both indices can be misleading; a more detailed energy performance model yields far greater insight into energy performance and overcomes the potential difficulties that accompany the use of energy intensity values.

This entry offers some working language for energy accounting, identifies the kinds of data that should be collected and analyzed, describes a number of approaches to energy performance assessment, and develops the basic principles and techniques of Monitoring and Targeting as the recommended approach to energy accounting.

Keywords: Accounting; Energy performance indicators; Energy intensity; Specific energy consumption; Degree days; Monitoring and targeting.

OVERVIEW OF ENERGY ACCOUNTING

Defining Energy Accounting

The adage, "if you don't measure it, you can't manage it" clearly applies to energy use. Just as financial accounting is necessary for effective management of an organization, energy accounting is a key element of energy management. Simply put, energy accounting is a system to measure, record, analyze, and report energy consumption and cost on a regular basis.[1]

Energy accounting systems typically consist of three parts: (1) a system to routinely monitor energy consumption and the variables that influence consumption, (2) an energy use record and reporting system, and (3) a performance measure.[2]

The effectiveness of the energy accounting system depends on the rigor with which consumption patterns are analyzed and correlated to independent variables such as production and weather. In many organizations, the energy accounting process is integrated with a statistical analysis methodology often referred to as Monitoring and Targeting. This is addressed in more detail in "Energy Monitoring, Targeting, and Reporting".

Depending on the goals of the organization, the accounting system may achieve some or all of the following objectives:

1. Track, record, and attribute energy consumption and costs.
2. Verify energy billings and troubleshoot errors.
3. Provide a basis for prioritizing energy capital investments.

4. Provide a basis for energy budgets as part of the overall budgeting process.
5. Identify unaccounted for energy waste.
6. Identify opportunities for performance improvement and evaluate the impact of performance improvement measures.
7. Optimize energy purchase practices.[3]

Energy Accounting and Energy Management

While many definitions of energy management are used, one that captures the essence of this organizational activity is:

> The judicious and effective use of energy to maximize profits (minimize costs) and enhance competitive positions.[4]

Energy management is consistent with other dimensions of continuous improvement. Key functions that comprise energy management include:

- Purchase or supply energy at the lowest possible cost.
- Ensure that energy is used at the highest possible efficiency.
- Utilize the most appropriate technology—from a business case perspective—to meet organizational needs.

Energy management is, or should be, an integral part of overall organizational management, and is practised by most of today's leading organizations in all sectors. Investments in energy management are generally sound, offering attractive returns to plant and building owners.[5]

Energy accounting is one of the tools employed, along with a variety of others, including policy and planning, training, communicating, investment appraisal, and operations and maintenance.

Methods of Energy Accounting

The methods employed in energy accounting depend on the nature of the organization's facilities, e.g., industrial plant or commercial building, and its objectives for the accounting program. In all cases, however, comparison of energy performance over time, and perhaps from site to site or building to building, is a likely element of the analysis.

A critical challenge when comparison of energy consumption is being carried out is the need to adjust consumption data:

- For changes in weather—heating or cooling degree days—over time or from place to place.
- For varying levels of activity, e.g., production or occupancy.
- For changes in space utilization in the facility, e.g., changes in conditioned floor space in a building.

Ultimately the methods described in "Energy Monitoring, Targeting, and Reporting" provide a rigorous statistical basis for these adjustments, but in many cases, other approaches are employed for performance comparisons. They include:

- *Present-to-past comparison*, in which energy consumption for a given period, i.e., a specific month, quarter or year, is compared to the same period in a previous or base year. Since there is no attempt to adjust for changes in weather from year to year, this comparison is rough at best, especially when applied to heating and cooling loads.
- *Multiple year monthly average*, in which the base for comparison is the average consumption of several years for the period in question; again, no adjustment is made for changes in weather, but the assumption is that variation in weather is eliminated as a factor by averaging several years (a flawed assumption if climate change patterns result in a trend in weather rather than random variation).
- *Heating/cooling degree day (HDD and CDD) adjustment*, in which average temperature and degree-day data are used to adjust heating and cooling energy consumption to a common base for year to year or period to period comparison.
- *Correction for changing conditioned floor area*, in which it is assumed that energy consumption will increase or decrease proportionately to increased or decreased conditioned area.[6]
- *Adjustment for changing production*, in which energy intensity (energy consumed per unit of production) is used to scale consumption up or down for comparison. While it is necessary to adjust for production, this approach is not recommended for reasons that are addressed in "Problems with Energy Performance Indicators".

Energy Accounting Tools

A commonly held view is that significant expenditures in metering, data collection systems, and software applications must be made before energy accounting can be done. However, many organizations have discovered that manual collection of data from existing meters and records, and manual analysis using methods such as those described in "Energy monitoring, Targeting, and Reporting" can yield useful results.

Manual energy accounting can be greatly facilitated with the use of commercial spreadsheet programs to automate the numerous calculations that may be required, such as energy intensity and energy consumption per square foot per HDD. As well, other embedded functions such as graphing, regression, averaging, and others, can be helpful for analysis and presentation of results.

Energy Accounting Software

As the energy accounting system becomes more sophisticated and complex, or in the case of larger or multi-site organizations, commercial energy accounting software packages are available. These packages make it easier to input data, carry out analysis, and generate reports. They also typically incorporate weather and floor area corrections, and may enable the direct download of energy consumption, demand, and cost data from service meters or utility-based web sites.[7]

A number of accounting software packages are available commercially. The following list provides some examples (no effort has been taken to ensure that this list is complete or to verify the functionality of the packages):

- Envision™, a stand-alone hardware and software package for tracking energy use; Energard Corporation, Redmond, WA, www.energard.com
- Faser™, software for tracking, analyzing, and reporting utility bill data; OmniComp Inc., Houston, TX, www.omni-comp.com
- Metrix™, an energy accounting system that focuses on energy projects and the savings related to the projects; SRC Systems Inc., Berkeley, CA, www.src-systems.com
- Utility Manager™, software that targets the commercial and public sector markets, including school districts and local governments; Illinova Energy Partners, Oak Brook, IL, www.illinova.com/iep.nsf/web/IEPSubsidiaryHome[8]
- Meter Manager™, a system that combines a utilities supervision function with sub-metering and aggregated metering; Carma Industries, Peterborough, ON, www.carmaindustries.com
- Global Mvo Asset Manager™, a multi-purpose software package combining metering and sub-metering technologies with on-site and remote measuring, verification, and operational capabilities; Global Facman Enterprises Inc., 12180 Chemin du Golf, Montreal, QC, www.globalmvo.com[9]

STRUCTURAL CONSIDERATIONS—ENERGY ACCOUNT CENTERS

Organizations typically identify cost centers for financial management. For similar reasons, including the assignment of accountability to line managers, energy account centers (EAC) are helpful in organizing for energy accounting.

Energy account centers work by identifying geographically definable areas of management accountability, installing meters on energy utilities, and energy-based utilities (e.g., steam from a central plant, compressed air, etc.) at the point of entry to the EAC department or operating unit, and providing consumption reports as a component of the management information system.[10]

There are constraints and guidelines that apply to the selection of energy account centers that do not apply necessarily to financial cost centers. If possible, selection should be based on the following criteria:

- If sub-metering is required, the potential cost savings from energy reduction justify the cost of installing new metering.
- Energy consumption can be measured.
- Ownership of the energy account center can be established. The center might be a production department, a single meter, an aggregate of several meters, or other possibilities.
- An activity variable is identifiable. In the industrial sector, the variable may be production associated with a production unit that is established as an energy account center, whereas in certain building sectors, the variable may be occupancy rates (more about this in "Energy Monitoring, Targeting, and Reporting").
- There is a linkage between the account center and the organizational structure. Accountability can be assigned to an appropriate manager, and reporting can be integrated fully in the management information system.[11]

KEY ACTIVITIES IN ENERGY ACCOUNTING

Tabulation of Data

Energy consumption data is available from accounting records. Utility and fuel supplier invoices contain valuable information about consumption that can be tabulated. For various reasons, fuel and electricity consumption data must be treated separately.

For electricity, the following data should be collected and tabulated: billing month, number of days in the billing period, demand, and energy. From these data, a number of derived factors should be determined and tabulated, including daily energy, energy cost as a percentage of the total, demand cost as a percentage of the total, total cost, blended or average cost per kWh, and load factor as defined by Eq. 1:

$$\text{Load factor}\,(\%) = \frac{\text{kWh used in period}}{\text{Peak kW} \times 24\,\text{h per day} \times \#\,\text{days in period}} \times 100 \qquad (1)$$

For fuels, data should be recorded in physically measurable units (cubic feet, gallons, etc.) rather than dollars that can fluctuate over time (e.g., via utility rate changes, product price changes). Where two different energy sources feed thermal energy data into the same system, it may be necessary to convert them to a common unit. In a spreadsheet program, units can be converted as needed after the quantities are entered in their original units.

In addition to energy use data, data on the factors that influence energy usage are collected and tabulated, including production quantities, outside air temperature, time the facility is occupied, and so on.

Calculation of Energy Performance Indicators

A key component of energy accounting is the determination of energy performance indicators that enable comparison of energy efficiency over time, from site to site, and against appropriate benchmarks. A number of indicators are commonly used to relate consumption to measures of activity, weather factors, and facility size, for example Btu/unit of production, Btu/degree day, Btu/ft^2, or combinations of these. Others that may be useful include Btu/sales dollar, energy spend/sales or profit or value added, Btu/direct labor cost.[12]

Several of these indicators find common use in specific sectors, as indicated in the following paragraphs.

Energy Utilization Index[13]

The energy utilization index (EUI) is used in the building sector, usually based on annual energy consumption related to conditioned floor space. The basic factor in Btu/ft^2, kWh/m^2, or some other appropriate unit is normalized by adjusting for operating hours, weather, etc. for energy type, or for energy use (i.e., heat, lighting, air conditioning, etc.).

Normalized Performance Indicator[14]

The basic performance factor or EUI in Btu/ft^2, kWh/m^2, or some other appropriate unit is normalized by adjusting for operating hours, weather, etc. for energy type, or for energy use (i.e., heat, lighting, air conditioning, etc.). The normalized performance indicator (NPI) is used for comparison of buildings of similar type.

Specific Energy Consumption[15]

This indicator is an energy intensity used in industry to relate consumption to production, expressed, for example, as MBtu/ton, GJ/unit, or other appropriate units. While commonly used, there is a real danger that these indicators can be misleading. Variation in the specific energy consumption (SEC) may be due to economies of scale, production problems, weather, or other factors that are not related to energy management. As well, it is important to consider the fixed and variable components of energy consumption, as discussed in "Problems with Energy Performance Indicators".

Energy Balance

Just as financial accounting involves the reconciliation of revenues and expenses, so energy accounting can (and should) reconcile energy inputs and outputs. Secure in the First Law of Thermodynamics, the principle that all energy can be accounted for, since it cannot be created or destroyed, enables the energy manager to balance inputs and outputs.

Inputs are relatively easily calculated on the basis of purchased energy, although energy exchanges between the facility and the environment may also need to be included, such as air infiltration/exfiltration in buildings, solar gain, and so on. Methods exist for calculating these.

Outputs or end-uses require an energy load inventory—for both electrical and thermal loads. While time-consuming, the preparation of a load inventory involves:

- The counting and tabulation of electrical devices in the facility, including their nameplate or measured demand and energy consumption and times of operation.
- The measurement or calculation of all thermal loads, such as burners in boilers and furnaces, steam or hot water flow, ventilation air flow, fluids to drain, heat loss and heat gain through the facility envelope, and so on.
- The calculation of total consumption based on the electrical and thermal inventories.
- Finding either the electrical load or thermal load for systems with varying loads is not always easy. Actual measurements or simulations may be needed to find loads of motors, HVAC, and boiler systems.

Total energy consumed should balance with total energy purchased within reasonable limits of error. If that is not the case, an unaccounted for loss may be awaiting discovery.

ENERGY MONITORING, TARGETING, AND REPORTING

An important element of Energy Accounting is the determination of the functional relationships between consumption and the independent variables that drive consumption. While often viewed as an issue separate from energy accounting, Energy Monitoring, Targeting and Reporting (MT&R) is an analysis technique that yields these functional relationships. As noted below, it is also a technique that provides a sound basis for energy budgeting, which is clearly part of the accounting process.

Working Definitions

By definition, MT&R is the activity that uses information on energy consumption as a basis for control and managing consumption downward. The three component activities are distinct yet inter-related:

- *Monitoring* is the regular collection of information on energy use. Its purpose is to establish a basis of management control, to determine when and why energy consumption is deviating from an established pattern, and to serve as a basis for taking management action where necessary. Monitoring is essentially aimed at preserving an established pattern.
- *Targeting* is the identification of levels of energy consumption, which are desirable as a management objective to work toward.
- *Reporting* involves putting the management information generated from the Monitoring process in a form that enables ongoing control of energy use, the achievement of reduction targets, and the verification of savings.

Monitoring and Targeting have elements in common, and they share much of the same information. As a rule, however, Monitoring comes before Targeting, because without Monitoring, you cannot know precisely where you are starting from or decide if a target has been achieved. The Reporting phase not only supports management control, but also provides for accountability in the relationship between performance and targets.

Energy Monitoring, Targeting and Reporting is consistent with other continuous improvement techniques applied in organizations, and should be viewed as an ongoing, cyclical process.

Energy Monitoring

There are two essential steps in energy Monitoring; they are:

1. The determination of a functional relationship between consumption and the independent variables that drive consumption (or what can be termed an energy performance model), typically production in the manufacturing environment, weather and occupancy in the buildings sector, or combinations of these and other variables.
2. The comparison of actual consumption to that predicted by the energy performance model.

The Energy Performance Model

Various methods of developing an energy performance model are used. Often linear regression produces a useful model relating consumption to production in manufacturing, or consumption to degree-days in buildings. In other instances, multi-variant regression on both production and degree-days or other combinations of variables is required to generate a useful model.

Energy used in production processes typically heats, cools, changes the state of, or moves material. While it is impossible to generalize, as industrial processes are both complex and widely varied, a theoretical assessment of specific processes gives reason to expect that energy plotted against production will produce a straight line of the general form:

$$y = mx + c \tag{2}$$

where c, the intercept (and zero production energy consumption), and m, the slope, are empirical coefficients, characteristic of the system being analyzed.

In the case of heating and cooling loads in buildings, a theoretical relationship between energy and degree-days typically takes the form of Eq. 3:

$$H = (UA + C_P NV) \times \text{degree-days} + c \tag{3}$$

where:

- H is the heat added to or removed from the building per unit of time.
- U is the heat transfer coefficient of the building envelope, taking into account its components such as glazing, interior wall finish, insulation, exterior wall, etc.
- A is the external area of the building envelope.
- C_p is the specific heat of air.
- N is the number of air changes per unit of time.
- V is the volume of the building being ventilated.

U, A, C_p, N, and V are all characteristic constants of the building. Eq. 3 is the equation of a straight line when H is plotted against degree-days, having a slope $=(UA + C_p NV)$ and an intercept on the y-axis $= c$. This constant c is the 'no load' energy consumed, no matter the weather conditions, by such things as office equipment, the losses from the boiler, lighting, and people.

Fig. 1 illustrates a performance model obtained from a consumption–production regression. In this case, the energy consumed in MMBtu is equal to 2.0078 times the production in pounds plus a constant 64,966 MMBtu (the intercept of the regression line) for a baseline period that is considered to represent consistent performance. It is important to recognize that total consumption typically consists of at least these two components:

- A variable, production-dependent load.
- A constant, production-independent load.

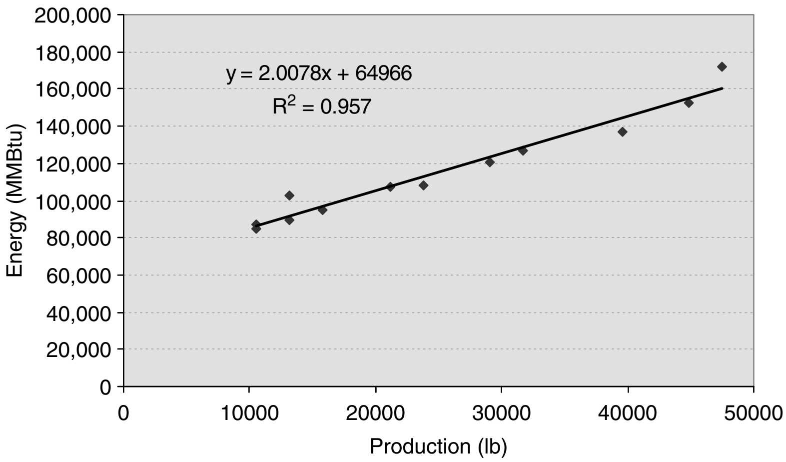


Fig. 1 Regression analysis of baseline for food processing plant energy consumption.

Similar models can be produced for buildings, in which case degree-days rather than production may be the independent variable.

In addition to providing a basis for the reduction of energy waste, the energy performance model also provides the means of determining the energy budget for a projected level of industrial activity or projected weather conditions in a future period.

Cumulative Sum

The comparison of actual and theoretical or predicted energy consumption uses is called cumulative sum of differences (CUSUM) analysis. In CUSUM analysis, a cumulative sum of the differences between the theoretical energy consumption calculated from the energy performance model and the actual consumption is calculated ($\sum$[theoretical − actual]). A time series plot of CUSUM values is illustrated in Fig. 2.

The CUSUM graph yields the following kinds of information:

- Changes in slope represent changes in energy performance.
- A downward (negative) slope represents consumption less than that predicted by the energy performance model, and vice versa.
- Cumulative energy savings or losses (in comparison to the energy performance model or baseline) at any point are equal to the ordinate of the point in question on the CUSUM curve.

Targeting

Based on the information derived from energy Monitoring, it is possible to set reduction targets in the form of energy performance models that:

- Represent best historical performance.
- Incorporate specified reductions to the fixed and variable components of total load.
- Eliminate periods of poor performance to establish a basis for future performance.
- Are defined by other similar criteria.

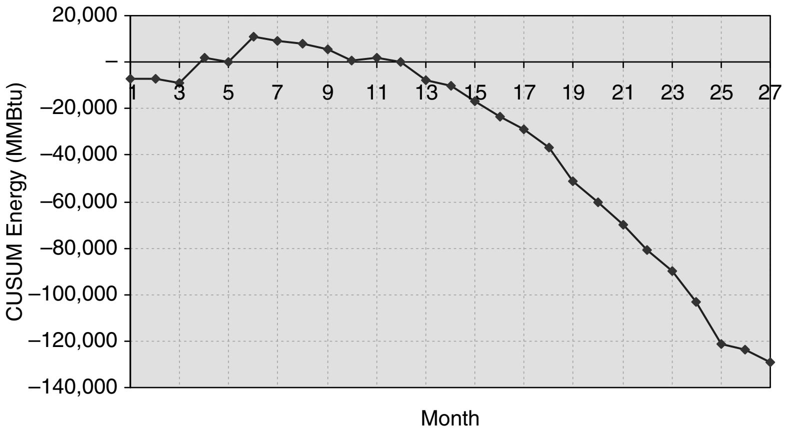


Fig. 2 Cumulative sum (CUSUM) graph for food processing plant.

Problems with Energy Performance Indicators

Especially in the industrial sector, energy performance is often expressed in terms of an energy intensity indicator, as discussed in “Energy Monitoing”. The energy performance model resulting from Monitoring analysis makes it evident that there are serious limitations to energy intensity indicators in providing an accurate measure of performance, as Fig. 3 illustrates.

As illustrated, it is possible to look at a performance point (1) in terms of its energy intensity, represented by the solid line, or as a point on the true energy performance model line represented as a dashed line. Points above the energy intensity line by definition have higher energy intensity values, and vice versa. Similarly, points above the energy performance model line represent worse energy efficiency, while points below represent improved energy efficiency.

If production were changed such that the performance point moved from (1) to (2), it would appear that energy intensity has decreased, that is, improved; however, quite the opposite is true when the real performance model is considered. Conversely, a decrease in production that changed the performance point from (1) to (3) would appear to worsen performance when indicated by the energy intensity, whereas, again, quite the opposite is true.[16]

It follows that the only reliable indicator of performance is that derived from the energy performance model; the simple, and widely used, energy intensity value must be viewed with real caution.

Reporting

Reporting within a Monitoring and Targeting system has a number of functions:

- To create motivation for energy saving actions.
- To report regularly on performance.
- To monitor overall utility costs.
- To monitor cost savings.

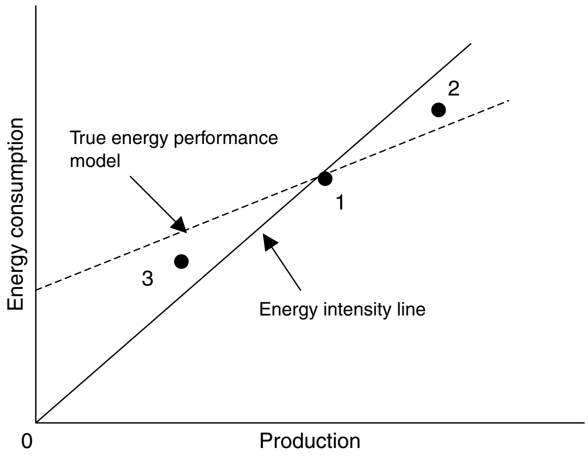


Fig. 3 The problem with energy intensity.

Within most organizations, the need for the type of information generated by a Monitoring and Targeting system varies with level and responsibility. Typically, as the need moves from the operational level in the plant to the senior management level, the requirement for detail diminishes, as does the frequency of Reporting. Operations staff need energy control information to stimulate specific energy savings actions. Senior managers need summary information with which to guide the organization’s energy management effort. One report for all will not result in actions being undertaken and decisions being made.[17]

CONCLUSION

Successful organizations include energy management in their management information systems. At the very least, they track consumption, identify and respond to trends, base energy purchase strategies on detailed knowledge of their consumption patterns, and determine performance indices for comparison to their own facilities over time and benchmarks for other similar facilities.

Since the commonly used energy performance indices tend to be energy intensities (i.e., consumption per unit of production, consumption per unit of floor space, consumption cost as a fraction of total product cost, and others) their usefulness in driving performance improvement may be limited.

Energy Monitoring and Targeting is an approach to performance assessment that provides a basis for managing energy consumption downwards. It is based on the determination of an energy performance model that takes into account the fixed and variable components of total consumption. Energy Monitoring and Targeting can, and should, be the basis for effective energy accounting.

REFERENCES

1. California Energy Commission. *Energy Accounting: A Key Tool in Managing Energy Costs*. January 2000; P400-00-001B, 1.
2. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 4th Ed.; The Fairmont Press, Inc.: Lilburn, GA, 2003; 24.
3. California Energy Commission. *Energy Accounting: A Key Tool in Managing Energy Costs*, January 2000, P400-00-001B; 2–3.
4. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 4th Ed.; The Fairmont Press, Inc.: Lilburn GA, 2003; 1.
5. Natural Resources Canada (formerly Energy, Mines and Resources Canada). *Energy Accounting*; Canadian

Government Publishing Centre, Ottawa, Canada, 1989; ISBN 0-662-14152-0, 1.

6. California Energy Commission, *Energy Accounting: A Key Tool in Managing Energy Costs*, January 2000, P400-00-001B; 13–14.
7. California Energy Commission, *Energy Accounting: A Key Tool in Managing Energy Costs*, January 2000, P400-00-001B; 15–17.
8. California Energy Commission, *Energy Accounting: A Key Tool in Managing Energy Costs*, January 2000, P400-00-001B; A-1–A-3.
9. Enbridge Gas Distribution. *Energy Accounting Guide for Energy Savings*, adapted with permission from California Energy Commission. *Energy Accounting: A Key Tool in Managing Energy Costs,* January 2000, Scarborough ON, Canada, www.enbridge.com; 23.
10. ETSU; Cheriton Technology Management Ltd *Good Practice Guide 112: Monitoring and Targeting in Large Companies*; Department of the Environment, Transport and the Regions: Oxfordshire, UK, 1998; 2.
11. ETSU, The John Pooley Consultancy *Good Practice Guide 91: Monitoring and Targeting in Large Manufacturing Companies*; Department of the Environment, Transport and the Regions: Oxfordshire, UK, 1994; 8.
12. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 4th Ed.; The Fairmont Press, Inc.: Lilburn GA, 2003; 30.
13. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 4th Ed.; The Fairmont Press, Inc.: Lilburn GA, 2003; 26.
14. ETSU, The John Pooley Consultancy, *Good Practice Guide 91: Monitoring and Targeting in Large Manufacturing Companies*, Department of the Environment, Transport and the Regions: Oxfordshire UK; 1994; 14.
15. ETSU, The John Pooley Consultancy, *Good Practice Guide 91: Monitoring and Targeting in Large Manufacturing Companies*, Department of the Environment, Transport and the Regions: Oxfordshire UK; 1994; 14–15.
16. Stephen Dixon. *Private Discussion*, sdixon@knowenergy.com.
17. ETSU, The John Pooley Consultancy, *Good Practice Guide 91: Monitoring and Targeting in Large Manufacturing Companies*, Department of the Environment, Transport and the Regions: Oxfordshire UK; 1994; 20–22.

Air Emissions Reductions from Energy Efficiency

Bruce K. Colburn
EPS Capital Corp., Villanova, Pennsylvania, U.S.A.

Abstract
Energy efficiency has become a popular buzz phrase in the 21st century, but in addition to the pure economic cash savings that can be affected from implementing such work there is also the potential for a cleaner environment. Since becoming operational in February 2005, the Kyoto Treaty calls for a reduction in greenhouse gases (GHG) from all developed countries according to a rather strict time schedule. There is much debate as to whether those targets can be met through existing technologies. One of the obvious solutions to part of this problem is to dramatically increase energy efficiency programs because they permanently reduce the use of electricity and fuels. With the advent of the Kyoto Treaty, a trading mechanism for buying and selling CO_2 credits also can provide some organizations an additional financial incentive. In addition, there are various other undesirable emissions such as SO_x, unburned hydrocarbons, mercury, dioxins, and other undesirable "products of combustion" which are also reduced as a result of energy efficiency.

INTRODUCTION

Energy efficiency refers to the avoidance of waste in the utilization of energy, regardless of its source. Some electricity is not generated from fossil fuels but from hydro, wind, solar, biomass, and geothermal energy. All are considered a precious commodity, so whether fossil fuels or renewable sources are involved, the concept of being as efficient as possible in the use of energy in a facility or industrial process is a reasonable and business-like goal. In the process of evaluating energy efficiency options, a basic guideline is if electricity and thermal fuel usage is reduced, then, in general, a positive environmental condition will simultaneously occur. The simple mechanism of reducing electricity and fuel also reduces the resulting air pollution from most of these sources.[1–4] There is one exception: electricity from a renewable resource. But in practice, renewables currently supply only a small fraction of energy in most countries (Sweden is a notable exception). Regardless, a reduction in the use of energy will reduce the fossil fuels used somewhere in the system. This reduction in fossil fuels reduces the greenhouse gas (GHG) emissions and other products of combustion such as sulphur, mercury, and dioxins, amongst others.

Under the Kyoto Treaty, there are multiple means established for reducing air emissions in the six regulated gases: CO_2, CH_4, N_2O, hydro oxide, hydro fluorocarbons (HFC) and per fluorocarbons (PFC), and sulphur hexaflouride—SF_6. All have been configured in terms of CO_2 "equivalents" for purposes of calculating the avoided atmospheric environmental effects. The overall objective is substantial, permanent, worldwide reductions in total air emissions of CO_2 and other GHG equivalents. To do so, a series of techniques involving not only reductions within a country but the trading of "credits" from implementation by others in the same or other countries has been set up. Examples of these are Emission Trading (ET), Joint Implementation (JI), and Clean Development Mechanism (CDM), and are all focused on the major industrial countries (referred to as "Annex I" countries, which is similar to "Annex B" in the Kyoto Protocol) to implement reductions elsewhere by claiming credit for the business or government which made that reduction possible (hence the "trading" acronym). Similar in approach to the Kyoto Treaty, the United States set up the 1990 Clean Air Act Amendment set up a "cap and trade" system, similar to GHG, for SO_2 emissions.

The ratification of the Kyoto Treaty required that most industrialized countries (the United States not included) reduce their total GHG emissions. The opportunity arose for some organizations to sell these GHG reductions in units of metric tonnes of CO_2 to other organizations that, at the same time, feel they need them to "offset" what is otherwise a reduction requirement. In this way, a business can either invest in energy efficiency by purchasing equipment solely designed to sequester carbon and thus create reductions in net CO_2, or buy CO_2 "offsets" which will be recognized internationally and allow credits against their target reduction requirement. In all cases, capital expenditures of some form are required, but this range of options allows businesses to select the optimum mix. It is noted that although some persons may disagree that global warming exists, the data available suggests otherwise. Fig. 1 illustrates the dramatic rise in

Keywords: Energy efficiency; Emissions; Air emission; Greenhouse gases; GHG; Kyoto Treaty; Emissions trading; CO2 reduction; Clean air; Emission reduction.

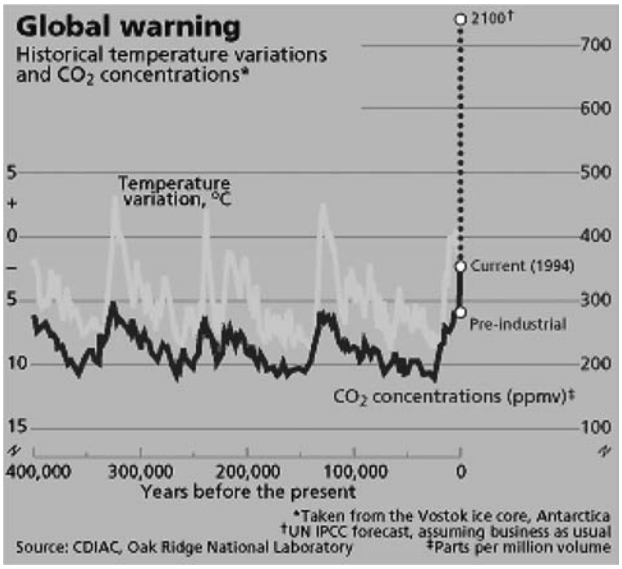


Fig. 1 Global temp and CO_2 vs time.

temperature as the CO_2 concentration increases while Fig. 2 shows the noticeable rise in both CO_2 and CO concentrations in the atmosphere.

Overall, energy cost reduction is the best solution for reducing GHG air emissions because it provides the only mechanism that allows a user to invest capital and reap a direct economic return on that investment while simultaneously receiving CO_2 reductions due to reductions in consumption of electricity and fossil fuels. All other approaches seem to rely on investing in capital, which is a burden cost—it does reduce the total GHG, but with little or no economic return (for example, a baghouse collecting particulate may have some trivial reuse value but it will be nowhere near the amortization investment and annual operating cost). Therefore, one of the best solutions is to invest in energy efficiency and, as a direct result, simultaneously improve the environment through a reduced need for energy.

Emissions measurements in metric tons ("tonnes") of CO_2 equivalents are the internationally accepted norm for

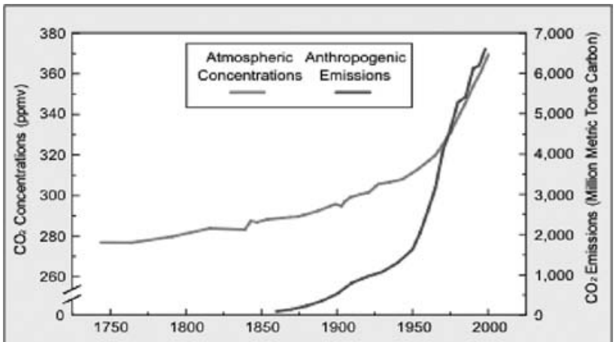


Fig. 2 CO_2 concentration vs time.

reducing the "global warming" problem. Various calculations have been made for GHG reductions, but most depend on such factors as the mix of electric generation from fossil fuels, which varies greatly around the world. For example, until the 21st century in Quebec, Canada, and Brazil, most electricity came from hydroelectric plants, which do not produce GHGs. There is a debate in the science community whether GHGs emitted from decomposing growing plants—which are destroyed when land is dammed up—are, in fact, a GHG "penalty" which should be charged to hydroelectric generation. Therefore, the argument follows that if one saves electricity consumption, one does not "help" the environment by reducing GHG emissions. Thus far, this argument is not widely accepted in the environmental community. However, in practice, all geographic areas have (inefficient, old) fossil fuel plants as "topping" systems at a minimum, to cover the electricity that could not be produced from base-loaded hydro or nuclear energy. In such cases, this energy efficiency would translate to a reduction in operation for these polluting fossil plants as the "last run" unit.

In the United States, decisions on energy efficiency tend to be made solely on the basis of pure economics. In some countries, culture, good sense, and government edict (not necessarily related to cost/benefit) have been the deciding criteria for the operational mix of generating plants. For example, the 'push' in Sweden in the 1930s for very well-insulated buildings has finally paid off financially. For decades, however, it meant only that people were more comfortable in the cold climate as compared to others in surrounding areas, regardless of the amount of utility bills paid. Despite a similarly cold climate, Norway does not possess the same well-insulated buildings as Sweden, uses about twice the energy per capita as its neighbour, and is only now pushing energy efficiency. This difference alone can be seen as cultural issues affecting energy efficiency in a given country and the use of fossil fuels to run the local economy. Japan imports about 97% of its energy, so it focuses strongly on energy efficiency in every aspect of life because it is economically prudent both for business and government to do so. On the opposite spectrum, the United States wastes the most energy and has historically done so for over 100 years, primarily due to historically low energy prices. Because energy was cheap, methods for energy reduction did not find much of a foothold in the United States until the 1970s. These examples illustrate that there are a variety of reasons for being energy-wasteful or energy-efficient, and cost effectiveness has not always entered that equation.

THE HYDROGEN ECONOMY ALTERNATIVE

The previously mentioned issues of emissions and the GHG problem of excessive CO_2 generation are greatly reduced if the input fossil fuel is pure hydrogen, H_2. Hydrogen is clean combustion (effectively no CO_2 formation) compared to natural gas (let alone other fossil fuels) and the result of combustion is almost pure water vapor. This is because when pure hydrogen is burned in the atmosphere, there is not an accompanying carbon molecule to recombine. Issues of CO_2 tend not to be present to the same extent (although there is atmospheric recombination). Therefore, scientists have noted that if automobiles ran on hydrogen, much of the current air pollution generated by transportation would be reduced, as would the GHG emissions of CO_2. However, very little hydrogen exists naturally in nature. Most of it is produced through industrial process means such as stripping hydrogen off of methane (CH_4) or using electrolysis to separate water (H_2O) into pure oxygen and pure hydrogen (which requires substantial energy). As the entire oceans are full of hydrogen (water is H_2O), it would be great if cost-effective methods of separation were available for this "infinite" potential energy source other than pure electrolysis, which is simple and very energy-inefficient.

The greatest benefit of hydrogen use could be in transportation, namely cars and trucks. However, there are serious infrastructure problems at the present time that currently limit the practical use of such technology. In conjunction with the cost of the fuel and lack of refueling systems, few vehicles would be capable of combusting hydrogen. Also, there is very little refining capability in producing large amounts of hydrogen and there is a lack of means to safely transport it around the country. One great benefit of hydrogen is its energy storage capacity in a small space, just as with natural gas or oil. In contrast, electricity is normally stored via chemical batteries or esoteric solutions as pumped hydro storage or compressed air caverns, which are large, very localized, expensive, and relatively inefficient. It is interesting to note, however, that small packages of compressed air storage for electricity is now available (2006) in modules that can fit in a closet and produce UPS power for personal computers and other critical power functions.

Above all else, it should be understood that hydrogen is not available in its natural state in meaningful amounts, and therefore, it virtually always requires energy to "produce" it. In practice, hydrogen is merely a method for "storing" energy—possibly convenient and with high energy density—but an entire infrastructure system of storage, distribution, and fueling would certainly have to be developed. Currently there is no structure, to say nothing of the production plant infrastructure worldwide which is very insignificant at this time. So when it is said that hydrogen combustion produces no emissions, it should also be understood that generating hydrogen causes GHG emissions, unless renewable energy sources are involved. So there may be no net effective GHG emissions reductions using hydrogen as a fuel source, despite the rhetoric of some. This total energy accounting is currently controversial because of arguments that only renewables such as solar and wind can be used to generate hydrogen. However, such arguments forget that without generating the hydrogen, those same renewables could have been used to displace existing fossil fuel production in the first place.

EMISSIONS IMPACTS

In most countries, there are now environmental agencies that oversee the regulation of pollution within its borders. In the United States, the Environmental Protection Agency has that charge and allows individual states to administer their own more severe rules (such as California). In most cases, equipment that directly combusts fossil fuels must secure air permits and have "permission to pollute," which usually involves having limits on the peak rate of generation of certain pollutants as well as total annual pollutants. In cases where the pollution would otherwise be too severe, scrubber systems as "control" devices are mandated to remove sulphur, mercury, or other undesirable products which are deemed unhealthy to the public. However, no country had ever regulated the amount of CO_2 emitted until the Kyoto Treaty. Systems have been developed which can either reduce or capture NO_x (nitrogen oxides, which cause smog), but none have been tested in commercial use yet for pure capture of NO_x or CO_2. With the modern understanding that GHGs are promoting global warming, steps are now underway in most countries to change that paradigm and drastically reduce the CO_2 emissions into the atmosphere through a combination of public awareness and new laws governing air emissions.

In Table 1 (in which all tonnes are metric), a typical set of emissions factors are shown, representing the weight of pollutants per MMBtu in a given category emitted from combustion of three common fossil fuels used today. These fuels are natural gas, #2 Fuel Oil (a form of light diesel), and coal. These emissions figures by themselves seem fairly tame at per million Btu [Higher Heating Value (HHV)], but when applied to a typical large industrial boiler they can lead to a lot of magnitude pollution per year. The second half of the table shows the impact of fossil fuel combustion[5] on some of the critical GHGs as a function of the fossil fuel source. It can be seen that merely performing a fuel switch from coal to natural gas can have a dramatic impact on the CO_2 generated for the same energy output result. This is one of the reasons that natural gas has become so popular as a fuel source.

In Britain and elsewhere in Europe, a new rigid agenda is forthcoming to large industries called the "carbon crunch," namely butting up against the allowable limits of CO_2 emissions for a given nation regardless of its population, economic growth, wealth, or any other parameters. This will soon mean that some corporations

Table 1 Air emissions from burning fossil fuels

Pollutant fuel	CO_2	CO	NO_x	Lead	PM10	Voc	SO_x
			Units of pounds/MMBTU fuel combusted				
Nat gas	121.22	0.082	0.03	0.0000005	0.0076	0.0055	0.0006
#2 Oil	178.4	0.04	0.169	0.0000005	0.016	0.008896	0.2603
Coal	240.5		0.6				1.75
Using 2,000,000 MMBTU natural gas, or #2 fuel oil, we obtain the following:							
			Metric tonnes pollutants for 2,000,000 MMBTU combusted				
Nat gas	110,000	74	27	0.0	6.9	5.0	0.5
#2 Oil	161,887	36	153	0.0	14.5	8.1	236.2
Coal	218,240		544				1588.0

will find that if they want to expand production, they will either have to implement major energy efficiency upgrade programs on a scale not seen previously or they will have to pay dearly for CO_2 reductions implemented by them and pay someone for credits through a free market where the highest bidder wins. Recent data suggests that in Britain, a tonnes of CO_2 credit may go for about 20 €, or about $25 USD, which is about five times the estimated value a few years ago—and the carbon crunch has yet to really begin! Not only are CO_2 emissions being capped by country, but they are being lowered—in the next 10 years, the total CO_2 emissions for most Kyoto signators will have to be 10% below their 1990 levels of CO_2 emissions. A notable exception is Russia, which, due to the collapse of the Soviet Empire and the retraction of industry, already has had a reduction in fossil fuel due to small scale "depression" from business downturn while the country tries to adjust to a free market economy.

Another factor in environmental emissions is the effect on marginal GHG emissions from generation due to the implementation of energy efficiency measures. The amount of CO_2 saved varies by region, season, and time of day (TOD) simply due to the mix of current electric generation equipment in place. Some states are attempting to receive credit for the energy efficiency impacts in their State Implementation Plans (SIPs) submitted to the EPA concerning electric power generation; this involves developing local-specific factors for the kWh savings and when they occur (day/night/weekend). At an earlier point, the Alberta Interconnected Systems (AIS) in Canada had their marginal generation emission rate at 0.211 T/MWh, but their average was 0.279 T/MWh. This shows the strange effect that the electrical energy efficiency impact has on their more GHG-friendly plants (it is likely that the base-loaded coal plants were unaffected). Some energy efficiency measures will have very different emissions profiles than others even within the same plant or building simply due to the time impacts of those savings. Examples are lighting, which saves more at night than in the day, and a chiller savings measure, which is likely to save more in the daytime than in the nighttime. In fact, normal energy efficiency savings off-peak have a lower economic value to a project than savings during the daytime hours. However, the emissions reductions impact at night, per MWh, may be larger than by day, because many electric utilities are base-loaded with coal and may actually reduce coal consumption at night. Those reductions by day may reduce only gas-fired equipment consumption. This further demonstrates part of the problem with GHG emissions: at present, there is a reverse effect in that the least cash cost savings can occur in such a way that it discourages saving measures that would otherwise provide the maximum GHG emissions reductions.

ENERGY EFFICIENCY AND ENVIRONMENTAL STEWARDSHIP COMPLEMENT EACH OTHER

No matter how you look at it, permanently reducing the volume of fuels and kWh used reduces the total raw fuel inputs as reducing fossil fuel combustion ultimately reduces air pollution. Differing mixes of fuels occur in different regions of a country for electric power generation, but every country uses electric power that has a meaningful fraction generated from fossil fuels—typically natural gas, oil, and coal. These three fossil fuels cause major air pollution, regardless of the country's current government's environmental regulations on clean combustion or scrubber technology.

The environmental aspects associated with energy efficiency can certainly assist in the reduction of pollution, but it will not be eliminated solely by aggressive energy efficiency.

Virtually every responsible scientist now concedes that the global warming problem is real, and the arguments now tend to be focused on how much human intervention can impact the reduction in global warming and at what rate. In July 2005, the Wisconsin Public Service Commission of the United States (the state regulating body for electric utilities) redefined and allowed energy efficiency investments by the electric utilities to equal footing with investments in generation, transmission, and distribution. This is apparently the first utility-regulating body to recognize legally that reducing kWh or kW has just as much economic value as investments in generating, transmission, and distribution equipment. This provides some rational economic basis for not only allowing but encouraging real energy efficiency by promoting less energy use during a time in which energy use is otherwise growing. Such encouragement, at the very least, could flatten out electric usage over time and allow the utility to reap a return on their investment at their normal return rates. Instead of rewarding a utility for investments in growth, it is essentially rewarding a utility for slowing or stopping growth in electric energy consumption. Several states require utilities to produce integrated resource plans; for example, PacifCorp has agreed to procure all available cost-effective energy efficiency resources before, and in addition to, conventional generation assets. California has a similar provision so as to promote energy efficiency in the electrical generation area. Many other states have alternate programs, but many are geared at attempting to promote some form of energy-efficient and alternate energy solutions within the bounds of cost effectiveness and still be reasonably acceptable by the public at large. This demonstrates that here in the United States, above all else, the states recognize and are responding to public desires, whereas currently the federal government still seems to be considering whether GHG emissions cause global warming. A coalition of north eastern states has banded together to create legislation referred to as "mini-Kyoto," which recognizes and accepts that ignoring the problem of global warming is no longer appropriate. Because the United States federal government has not and will not be leading the charge, those selected states will create their own more rigid rules even though it might initially cause some negative economic impact in those isolated states. The feedback, however, is that over time they will gain advantage by being a leader in the world, regardless of federal guidelines, and their individual states will actually benefit from the improved situation *vis a vis* GHG control.

EU countries have adopted the Kyoto Treaty as well as developed programs and incentives to promote both energy efficiency and emissions limits—it is clear the two are strongly related. Because all EU countries signed the Kyoto Treaty, there is also unanimity of purpose and focus in the EU. Because the United States national government has not ratified the Kyoto Treaty, numerous states are now taking the lead and imposing Kyoto-like rules and regulations out of a common sense of need and purpose. Fig. 3 shows the huge disparity between total CO_2 emissions by the United States and other countries. This is further clarified by Fig. 4, showing the per capita emissions of CO_2 per country (per country emissions data from UN Millennium Indicators, 2005). In both cases, the United States is the biggest offender of any country by far, and when compared to Europe, is downright profligate (the EU as a whole is less than half the CO_2 per person than the United States or Canada and can be easily traced to cheap energy in North America). The development of the Chicago Climate Exchange is another example of a voluntary program in the United States that has begun without government initiative. Although trading is slow, it is increasing. More companies, cities, and states in the United States are joining to demonstrate their commitment to reducing global warming, even if it costs them and their citizens some additional money in the short-term. The nightmare scenario, however, is for fifty U.S. states to adopt fifty sets of laws that all differ substantially. Remember that California alone has the world's sixth largest economy, so some of these state programs may have tremendous influence in Congress by forcing a

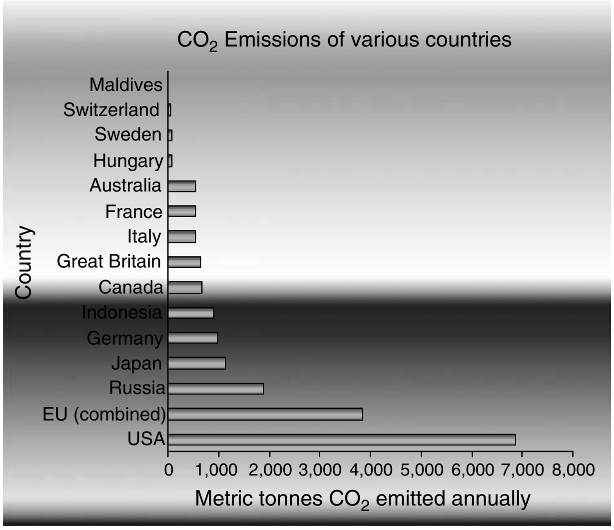


Fig. 3 Annual metric tonnes CO_2 by country-1998.

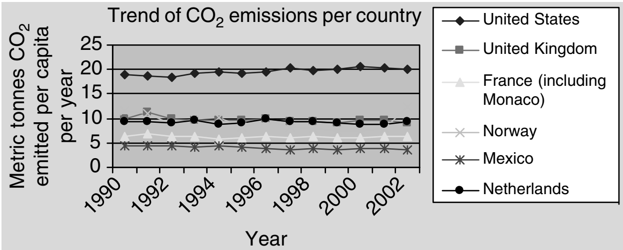


Fig. 4 Per capita CO_2 emissions by country-trends of CO_2 emissions per country.

rational national plan despite of the United State's refusal to ratify the Kyoto Treaty.

In the United States, a program with RECs (Renewable Energy Credits) is one way to reap additional financial benefit out of a renewable electrical energy production project—the RECs can be sold as certificates proving that the recipient has obtained renewable (and hence GHG-free) generation. The state of Pennsylvania has gone even further and identified energy efficiency results in terms of an "Alternate Energy Credit" (AEC) source, which can be sold (similar to the idea that elimination of a Watt-hour is a "negawatthour"). These revenues can thus be initially used to help support the energy efficiency investment and reduce environmental pollution and GHGs. These RECs and AECs are all based on 1 MWh electricity "generation" units. Depending on the location of the source of the electrical power generated, 1 MWh could represent 0.3–0.5 T of CO_2 avoided. Prices for these RECs vary by state or region, depending on the base cost of distributed electricity in a given area. The new AECs are predicted to be as low as $25/MWh in Pennsylvania due to lower-cost power, while prices in Massachusetts have been about $55/MWh. This REC credit is in addition to the actual raw KWH sale itself.

CARBON SEQUESTRATION

One area of renewables which has created some controversy is that of biomass—namely the use of grown crops as a fuel source. Such crops absorb CO_2, much as humans breathe oxygen for survival. The concept for emissions credit is that if one combusts biomass, then the CO_2 returned to the atmosphere will be absorbed by other crops, which will then be harvested and consumed as fuel. Similarly, forests and other greenery are considered "sinks" for CO_2, and the focused process of creating such sinks is referred to as "carbon sequestration" (literally grabbing CO_2 from the atmosphere). The concept is that by using renewable crops in boilers, one can generate electricity (or thermal heat for steam for processes) and the CO_2 emitted is actually the "gas" needed by the next round of crops for their breathing source. This creates a mutually circulating fluid wherein the generated CO_2 actually helps support the absorption of CO_2, which helps both industry and the environment. As part of any such technology, appropriate equipment must be used to scrub the exhaust for undesirable particulates and sulphur compounds and to control nitrogen compound discharges. The process, demonstrated in Fig. 5, is not perfect, but from a pure CO_2 viewpoint it is considered "neutral".

Carbon sequestration has led to some major research, including technologies and techniques for capturing the carbon before it enters the atmosphere after fossil fuels or other fuels are combusted and methods to sequester the carbon captured in the ocean.

The practical technologies, some of which exist in a fairly efficient (but not cost-effective) way, include solvent absorption/scrubbing; physical adsorption; gas absorption membrane systems (which themselves require a lot of energy to operate effectively); cryogenic fractionation (supercooling and distilling liquid CO_2 for separation and then some form of disposal as a concentrate); and chemical looping (in which flue exhaust gases are contacted with special metal oxides, which release oxygen for combustion but capture the carbon). By early 2006, a consortium of BP and Scottish & Southern Energy intend to have on-line a new power plant that strips the hydrogen off natural gas

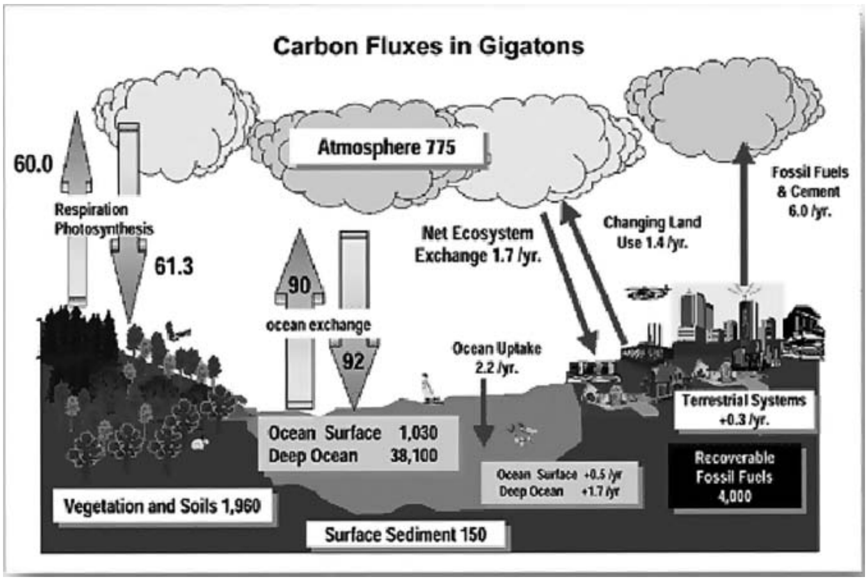


Fig. 5 Carbon sequestration options. (Courtesy US National Energy Technology Laboratory).

and captures the resulting CO_2 at the source before it can be emitted to the atmosphere. This CO_2 would be sequestered as a solid material and then injected into the ocean bottom, and the resulting H_2 would be using for fuel to generate electricity without generation of CO_2 in the exhaust. This would be the first industrial scale demonstration of sequestration. The most common method available presently is to plant more forest area and let the carbon dioxide released from normal combustion be absorbed over time into the growing tree plants, which then produce oxygen from photosynthesis as the trees grow. This "offset mechanism" approach still has the negative effect in that, initially, the CO_2 is discharged first into the atmosphere and only over time is a portion of it absorbed. However, this approach is inefficient because it also requires large plots of land—unless the forested land can grow harvestable cash crops that can be reused on a continuing basis. This is the new focus of some biomass power generation products.

EXAMPLE OF ENERGY EFFICIENCY YIELDING ENVIRONMENTAL BENEFITS AND COST EFFICIENCY

Consider using sludge in an anaerobic digester at a municipal wastewater treatment plant (WWTP) in India simultaneously to cogenerate electricity and heat, use the heat to produce more biogas, more quickly destroy sludge solids (thereby reducing the sludge dewatering process and ultimately reducing the amount of waste solids to be carried away by trucks), and reduce the environmental pollution (which presently flares 100% of the biogas and produces the most emissions). The baseline production of biogas through anaerobic digesters was originally low only because no digester gas was used for heating the sludge via a hot water boiler, due to the mild weather conditions in-country and the original desire for low first-cost construction of such a plant. Controlling operating costs was not a priority originally, but instead the need was for rapidly installing and operating WWTPs where there had originally been none. However, this means that the unheated sludge circulating in the anaerobic digesters does not decompose very quickly, limiting the destruction of solids in the sludge and the amount of gas generated (which then flares). The unheated sludge's average temperature was 24°C, whereas it typically heats to 36°C–40°C for optimum decomposition (36°C being considered optimum due to tradeoffs in energy required to heat the material and biological activity). The plant electrical load was approximately 850 KW (peaks about 950 KW), all from the electric grid, with TOD utility rates.

An ESCO project utilizing cogeneration, peak shaving, and ultrasound to improve the economics was planned for this (unheated) anaerobic digestion process. The energy measure consisted of 750 KW (BF?) cogeneration, a 250 KW peak shaving generator (the actual operation of the peak shaver depends on the actual peak demand during the on-peak electric period), heat recovery for heating the digesters, and an ultrasound system for further breaking down the fibrous materials in the sludge, thus generating more biogas due to additional surface area exposed in the sludge. The end result of such an approach was substantially greater biogas generation on-site and a further dewatering of the sludge and destruction of volatiles. Part of the reason for such a project was the desire of the government to promote energy efficiency (thus lowering operating costs) and demonstrating CO_2 offset benefits (because the flared gas would be used in offsetting other electrical generation). In addition, the resulting sludge was originally pressed and dewatering was accomplished with a fuel oil-fired dryer on-site to further remove water before the final sludge was hauled by truck to a landfill (and with the gas flared in sight of the fuel oil-fired dryer). This trucking also used fossil fuel, and the landfill consumed fuel oil for equipment to distribute the material around. The project reduced the amount of dewatering required and thus decreased the remaining sludge weight slightly by producing the additional biogas through further decomposition. This resulted in even further (secondary) CO_2 reductions through less trucking and landfill activity.

Based on the engineering analysis performed, the total installation cost was determined to be about $2,100,000. Because the existing biogas was naturally generated without heating, the resulting products from the plant was dryer sludge that would have further decomposed at a landfill, letting off additional unconstrained CH_4 (with an ozone damage about 21 times that of the CO_2 generated from combustion). So, not only was there the CO_2 generated from the flare of gas, but also the unburned hydrocarbons given off from decomposition. Being able to generate more gas on-site and using it on-site to offset grid electricity generated by a cogeneration plant does not produce more net CO_2 at the plant site. This is because, in this case, the CH_4 (and associated unburned methane from material decomposition) not generated from combustion under the project would be naturally generated over time through decomposition, but with no accompanying environmental or financial benefits. For illustrative purposes, only the equivalents for NO_x and SO_x are ignored herein. Table 2 shows the results of an energy and environmental efficiency project which also generates meaningful GHG reductions as a direct part of the project, and additional economic enhancements possible due to this environmental benefit on what otherwise might be considered only an energy efficiency project. These economic enhancements include CO_2 credits that could be sold to the EU or elsewhere to buy down the cost of the project.

This information shows how the upfront purchase of future CO_2 credits[6] for avoided CO_2 can be beneficial in

Table 2 Benefits of CO_2 credits for environmental projects

Wastewater treatment plant (WWTP) biogas enhancement and recovery in India (MCF = 1000 cubic ft at standard conditions)		
Life of project, years	15	
Original bio gas generated/year, MCF/year	137,664	
Bio gas generated/year after Project MCF/year	137,664	
Percent methane in biogas	65.0%	
BTU content biogas, BTU/ft^3 HHV	650	
KWH/year generated on site before retrofit	0	
Cogen heat rate (assumes all auxiliaries within), BTU/KWH HHV	15,000	
KWH/year generated on Site after retrofit	5,447,894	
Current grid elec producer cogen heat rate-nat gas, BTU/KWH HHV	8500	
Amount of methane gas for engines, MCF CH_4	89,482	
Amount of inert CO_2 gas, MCF CO_2	48,182	
Hours per year operation	8000	
Amount CO_2 generated by grid cogen plant, tonnes/year	2437	
Amount CO_2 avoided by trucks to landfill, tonnes/yr	128	
Amount CO_2 avoided by equipment at landfill, tonnes/yr	5	
Lifetime CO_2 avoided by Project	38,552	
Implementation cost, USD	$2,106,978	Already has digesters
Implementation cost. Rs	105,348,879	
Savings benefits/year, Rs	21,587,885	
Raw payback	4.9	Years-not attractive
Average cost/KWH, Rs./KWH	3.96	(U.S. 8 cents/KWH)
Necessary sale price to be financiable, Rs	86,351,540	Management Decision
Minimum buydown required for financing-U.S. company grant	18,997,339	Rs. $379,946.78
Cost/tonnes CO_2 avoided, Rs	493	Rs.
Cost/tonne CO_2 avoided, USD/tonne	$9.86	Minimum acceptable
Payback with 9.86/tonne Buyback of CO_2 credits	4.0	
Buydown at $30/tonne	$1,156,559	E.U. Price in 8/05
Net price after applying carbon credits to sale price	$950,419	
Net payback with E.U. free market CO_2 prices 8/05	2.2	Years Very Attractive Investment

assisting the economics of a project. The difficulty in such a case as this is that the avoidance of kWh purchases becomes part of the "savings," but those savings come from elsewhere, even though, with strict measurement and verification protocols, the true benefits can be documented and certified. It also shows how a net sale price of $10/T of CO_2 could dramatically affect the payback of an industrial project in India, and yet that price is less than half of the currently expected price in the EU as of fall 2005. This demonstrates that utilizing the digester gas in a responsible manner cannot only reduce operating costs but also reduce what otherwise is pumped into the atmosphere as additional tonnes of CO_2. It further shows how an EU business investing in a country like India—where the GHG credit limits do not currently apply yet credit is received for these certified GHG credits in the EU—can be beneficial, especially when they could invest about 40% of the unit price and receive equal (per metric tonne) GHG reduction credits.

To put this in perspective with the 493 Rs/T of CO_2 credit and using the annual CO_2 savings and the economic impact for the stated kWh avoided/year, we calculate a net added GHG impact savings of 0.2326 Rs/kWh, which compared to 3.962611 Rs/kWh average price is only about 6% improvement in the annual payback calculation. However, by being able to sell upfront many future years of GHG benefits and then using that money to reduce the net capital cost, there is a dramatic impact on the economic viability of the ECM, going from about 4.9 years simple payout to 4.0 years, pulling the project into the minimum range for selection. So, considering that the GHG credit only created about a 6% increase in annual equivalent, the economic benefit is multiplied many times over, if one can plan for 15 years into the future of reliable operation. It is the long-term years of multiplier that create the dramatic impact as well as the fact that there are currently organizations willing to pay upfront for those future GHG credits on the basis that there is a valid M&V methodology in place for verifying the true future efficiency and operational status of the resulting ECM. If one considers a gas-fired plant today, with U.S. prices of about $15/MMBtu (HHV), the fuel value alone of a generated kWh could be as high as 0.15 USD, but the GHG value of avoiding that kWh at the U.K. price of about $30/T CO_2 is about 1.63 US cents/kWh, or only about 11% of the fuel price. That benefit is converted to a 15-year upfront stream of money and could dramatically cut the cost of the project by 30%–40% due to the upfront nature of capitalizing the future cash flow stream.

CONCLUSIONS

The quest for reductions in environmental pollution—whether it be different chemical compositions such as SO_x, CO_2, NO_x, etc. or whether it be simply heat dumped to the atmosphere (such as a cooling tower does)—shares a common thread in energy efficiency.[7–9] The simple reduction in energy consumption for doing the same job reduces the requirement for energy to be produced. Because the majority of electrical and thermal energy comes from fossil fuel combustion, it simply reduces the need for these fossil fuels in order to achieve the same result. If such efforts were to be accomplished on a vast international scale, it might be possible for the existing fossil fuel pricing scenarios to be meaningfully altered. Also, the reduction in GHGs only helps reduce the global warming situation while at the same time reducing the other air pollutants such as NO_x and SO_x. The beauty of the energy efficiency approach to environmental air pollution and GHG reductions is that the cash cost reductions, which can be achieved by selling the environmental credits in the appropriate market, can help pay for the simultaneous environmental improvements in many cases. The ongoing work in carbon sequestration could find itself clashing somewhat with energy efficiency. If simple inexpensive sequestration were possible, the energy efficiency aspect of environmental pollution might not receive the same level of focus as it does currently due to pollution laws that mandate businesses to invest first in carbon sequestration, even to the exclusion of energy efficiency measures. However, this is unlikely to happen. Fortunately, there is mounting evidence that energy efficiency is growing in importance as a key element in the fight to reduce GHGs because in general it relies on proven technologies with long-term value.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Mr. Kevin Warren, P.E. of Warren Energy Engineering, LLC in Lincoln University, PA for his detailed and insightful comments and for the critique of the above essay. His input, without exception, was incorporated into the manuscript and the author believes it has added clarity and completeness to the topics (however limited in this document due to space constraints).

REFERENCES

1. Haberl, J.; Culp, C.; Yazdani, B.; et al. *Energy Efficiency/Renewable Energy Impact In the Texas Emissions Reduction Plan (TERP)*, Volume I-Summary Report, Energy Systems Laboratory; Texas A&M University: College Station, TX.
2. *Sector-Based Pollution Prevention: Toxic Reductions Through Energy Efficiency and Conservation Among Industrial Boilers*, July 2002, The Delta Institute, Chicago, IL. (http://www.delta-institute.org).
3. Colburn, B.; Walawalkar, R. *Global Emission Trading: Opportunities and Challenges for ESCOs*, Proceedings of the 2004 WEEC, Austin, TX, September 2004.
4. York, D. *Energy Efficiency and Emissions Trading: Experience from the Clean Air Act Amendments of 1990 for Using Energy Efficiency to Meet Air Pollution Regulations*, Report U034, American Council for an Energy-Efficient Economy: Washington, DC.
5. Mullet, M.R.; Schultz, S.C.; et al. *Wise Rules for Industrial Efficiency: A Tool Kit for Estimating Energy Savings and Greenhouse Gas Emissions Reductions*; U.S. Environmental Protection Agency Report EPA 231-R-98-014, July 1998.
6. New magazine called "Carbon Finance," available through http://www.carbon-financeonline.com.
7. Martin, N.; Anglani, N.; Einstein, D.; Khrushch, M.; Worrell, E.; Price, L.K. *Opportunities to Improve Energy Efficiency and Reduce Greenhouse Gas Emissions in the U.S. Pulp and Paper Industry*; Lawrence Berkeley National Laboratory Report LBNL-46141, July 2000.
8. *Clean Air Through Energy Efficiency*, 2004 Texas SB5 Report From the State Energy Conservation Office (SECO), State of Texas.
9. Pendergast, D. *Science and Technology Development to Integrate Energy Production and Greenhouse Gas Management*, Presented to Americas Nuclear Energy Symposium, October 2004, also http://www.computare.org/Support%20documents/Publications/Energy%20and%20CO2%20Management.htm.

Air Quality: Indoor Environment and Energy Efficiency

Shirley J. Hansen
Hansen Associates, Inc., Gig Harbor, Washington, U.S.A.

Abstract
In the days following the oil embargo of 1973, it became common practice to cover outside air intakes. This was just one of many actions taken by the uninformed in the hope of reducing energy consumption. Many of these measures, unfortunately, had a negative impact on the quality of the indoor air. Out of such ignorance came an assumption that energy efficiency (EE) and indoor air quality (IAQ) could not both be served in the same facility.

Over the years, the owner's dilemma regarding IAQ and EE has persisted. Many professional facility managers and real estate managers perceive only two options. There is the constant demand to run facilities as cost effectively as possible, which means that EE should be given a high priority. Unfortunately, many believe that this will result in poor IAQ, which can hurt productivity and/or lose tenants. They fear that a focus on IAQ will drive up their energy costs.

Today, we know that the IAQ risks associated with EE are more perceived than real. Yet fears remain that EE measures may have a negative impact on IAQ. These fears have increased the perception of IAQ risks, created EE sales resistance, and changed the financial dynamics of many projects. Recognizing that these fears exist and need to be treated is a critical first step in serving EE needs. This article addresses those fears and the real relationship between EE and IAQ.

INTRODUCTION

To examine the concerns related to energy efficiency (EE) and indoor air quality (IAQ), and to establish ways to achieve both in a given facility, it is important to:

1. Identify the sources that have linked EE and IAQ and determine whether any causal relationship between the two exists
2. Assess the advantages and disadvantages of ventilation as an IAQ mitigator
3. Consider ways that EE and IAQ might be compatible in a given facility

For years, the second or third paragraph of nearly every IAQ article has mentioned the energy crisis of the 1970s, the resulting tight buildings, and the growing IAQ problems. Readers have been left with the impression that as energy prices soared in the 1970s, owners and facility managers tightened buildings to save money and left occupants sealed in these tight boxes with pollution all about. These fears seem to be substantiated by a report by the National Institute of Safety and Health (NIOSH).

In its early report of investigations to date, NIOSH stated that 52% of the IAQ problems found were due to "inadequate ventilation." Somehow, that got translated to "inadequate outside air." A more careful look at that NIOSH's 52% figure reveals that such a translation misrepresented the findings. The "inadequate ventilation" problems encountered by NIOSH included

- Ventilation effectiveness (inadequate distribution)
- Poor HVAC maintenance
- Temperature and humidity complaints
- Filtration concerns
- Inappropriate energy conservation measures

Inadequate outside air was only one of a long list of problems

National Institute of Safety and Health also pointed out that the 52% figure was based on soft data. To the extent, however, that they represented primary problems in the investigated buildings, the NIOSH findings imparted another critical piece of information that typically is overlooked: of the problems NIOSH found, 48% were not solved by ventilation. National Institute of Safety and Health determined that nearly half of the problems it had investigated were not related to ventilation. If the NIOSH data and problems identified by other investigation teams are considered collectively, it seems safe to surmise that a great many of our indoor air problems cannot be satisfied solely by increasing outdoor air intake.

Somehow, indoor environment thought processes have been permeated by the idea that a tight building is not good and that it uses only recirculated air. Too often, ventilation has been perceived as being the preferred answer—which, of course, has increased energy consumption.

For nearly two decades, ventilation advocates have almost convinced facility managers and consultants that

Keywords: Indoor air quality; Tight-building syndrome; Sick-building syndrome; ASHRAE 62; Contaminants; IAQ mitigation.

opening the windows is the only measure needed for the air to get better "naturally." In the interest of both IAQ and EE, a careful look at a broader range of options is needed.

THE "FRESH-AIR" OPTION

If the air outside contains more contaminants than the inside air does, an outside-air solution may not be the answer. Fresh, natural air sounds wholesome, and it seems to be an attractive option. However, that natural air can be heavily polluted. When stepping outside the United Airlines terminal at O'Hare International Airport, for example, even a casual observer can tell that the air outside is much worse than the air inside. There is no "fresh air" for the O'Hare facility people to bring into the terminal. Natural ventilation could be a disaster. Opening the windows is not a viable option.

Hay-fever sufferers also tell us that opening the windows and letting in natural fresh air won't work. Between sniffles, they argue strongly against it.

From another perspective, we should analyze what happens inside when we open the window. What seemed like a good idea can cause a stack effect, in which warm air rises and pressure increases near the ceiling or roof. If we are concerned about a classroom, we could create negative pressure in the basement. Should that school have radon problems, cross-ventilation could cause even more radon to be drawn into the classrooms.

An alternative may be to induce outside air mechanically; that air then can be filtered and diffused through the facility. This method may be helpful, but it is not without problems.

VENTILATION CONSIDERATIONS

Ventilation is not always the answer. If we are to clear the air about the relationship between IAQ and EE, we need to make that statement even stronger. Ventilation is seldom the best answer. Certainly, it is an expensive answer.

The ASHRAE 62 standard is titled "Ventilation for Acceptable Indoor Air Quality." To the uninitiated, that sounds as though ventilation will deliver "acceptable" air. It may not. At the very least, the title implies that an organization as prestigious as the American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. has given its blessing to ventilation as the mitigating strategy.

As the various versions of ASHRAE 62 have been formulated over the years, the idea that most of our IAQ problems can be cured by ventilation has prevailed. ASHRAE 62 has, of course, brought relief to many, many people, who otherwise would have suffered from sick-building syndrome. The ASHRAE 62 standard met a key need during the years when it was very difficult to determine what some of the pollutants were, what their levels of concentration were (or should be), and what their sources were. Investigation and measurement protocols have come a long way since the first ASHRAE 62 standard was written, but we are not there yet. Increased ventilation can continue to give relief to occupants when we aren't quite sure what else to do.

Ventilation, however, is not the preferred treatment for IAQ problems—and it never has been. The U.S. Environmental Protection Agency has been telling us for years that the best mitigating strategy is control at the source.

In the 1970s and 1980s, outside air was reduced so that we wouldn't have to pay the higher energy costs of conditioning air and moving it around. With less outside air, we suddenly became aware of the contaminants that had been there all along. Less outside air meant greater concentrations. Because reduced ventilation was a fairly standard remedy in the 1970s, it is not surprising that the knee-jerk response to the air-quality dilemmas has been to increase ventilation.

Drawing more air into the building and blowing it around, however, has not necessarily solved IAQ problems. Sometimes, in fact, it has made things worse.

LOSING GROUND

Ventilation has created some IAQ problems where they did not previously exist. Two cases in point will help document the problems that the more-ventilation "remedy" fosters.

Relative Humidity

Historically, when construction costs exceeded the budget, one of the first ways to cut costs was to remove the humidifier/dehumidifier equipment from the specs. Today, without those humidifiers or dehumidifiers, it is very hard to correct the negative impact of increased ventilation on relative humidity. To reduce potential indoor pollutants where IAQ problems may not exist, increased ventilation has invited in all the IAQ problems associated with air that is too dry or too humid. With more than 50 years of data on respiratory irritation—even illness—due to dry air, creating drier air in colder climates suggests that we may be exacerbating the problem.

With all that is known about microbiological problems and their relationship to humid air, creating more-humid air in subtropical climates through increased ventilation is a questionable "remedy."

The Dilution Delusion

Increased ventilation thinking has prompted heavy reliance on dilution as the answer. Visualize, for a moment, all those airborne contaminants as a bright neon-orange liquid flowing out of a pipe in an occupied area. Would hosing it down each morning be considered to be a satisfactory solution? We have gained false confidence in dilution because the air pollutants cannot be seen—that does not mean they are less of a problem or that dilution is necessarily the solution.

The problem may not have been eliminated by reducing levels of concentration. There is still a lot that we do not know about chronic low-level exposure to some contaminants. A very real possibility exists that in a couple of decades, science may reveal that solution by dilution was nothing but delusion—a very serious delusion.

DETERMINING THE VALUE OF INCREASED OUTSIDE AIR

Using increased outdoor air as an IAQ mitigating strategy tends to make several other assumptions.

First, it assumes that increased outside air is going to reach the occupants in the building. As recently as the mid-1980s, a study of office ventilation effectiveness by Milt Mechler found that 50% of offices in the United States had ventilation designs that "short-circuited" the air flow. When considering possible treatment for IAQ problems, owners, EE consultants, and energy service companies should look at the facilities' air distribution system. Where are the diffusers? Increasing the outside air may cause a nice breeze across the ceiling, but it may do little for the occupants.

Second, the outdoor-air focus may prompt increased outside air when recirculated cleaned air may be better. Filtration and air cleaning were virtually ignored in ASHRAE's 62–89 standards and have not been treated sufficiently in subsequent work. Bringing in more outside air, which may be better than inside air, can still cost millions and millions of dollars. The fresh-air focus has too often overruled economics when specified filtration of recirculated air could provide the needed IAQ.

When unnecessary fossil fuels are burned to condition and circulate additional outside air, concern is raised about the impact on the quality of the outside air. A study conducted by the author that was reported at the Indoor Air 1991 conference in Helsinki determined that compliance with ASHRAE 62-1989 increased U.S. public schools' energy costs by approximately 20%. This measure not only expended a lot of precious tax revenue, but also offered an indication of the tons of additional pollutants that were put into the air each year.

POLLUTION SOURCES

Historically, the amount of outside air needed in a facility has been gauged by the CO_2 concentration in the air. This has been done because CO_2 is easier to measure than many contaminants are, so it serves as a good surrogate. Because people give off CO_2, it logically followed that the air changes per hour should be based on the number of people in an area. People-pollution thinking partially has its roots in the Dark Ages, when baths were not common and associated body odor was a major concern. Through the years, smoking problems have also led to using the number of people in an area as an air intake barometer. In fact, earlier ASHRAE ventilation standards were often referred to as odor standards.

Total reliance on people pollution has led us away from all the other pollution sources. We have subsequently had the Renaissance, the Industrial Age, the Technological Age, and the Information Age, each contributing new pollution concerns.

New volatile organic compounds (VOCs) are added to the list each year. As we "progress," people pollutants become less of a factor and building materials, furnishings, and "new and improved" equipment take on greater importance. Recent European studies have shown that the building pollutant load is much larger than we expected. When we measure our air intake per occupant, the pollutants created and dispersed to the outdoors by other sources are often overlooked.

Ventilation per occupant does not meet IAQ needs if pollutant sources other than people dominate an area. Laser printers and copiers, as they operate, give off just as many pollutants whether there are two people or 20 people in an office. Bioaerosols released from previously flooded carpet may pose as great a threat if there are 30 people in a room as they will with 300 people in the room. In fact, increased air circulation may draw air up from floor level and increase contaminants at nose level.

The problem has been aggravated by NIOSH's describing energy-efficient buildings as tight buildings. "Sick-building syndrome" and "tight-building syndrome" became synonymous. The idea became so pervasive that it prompted some very energy-inefficient operations. Too often, operable windows have been removed from building designs. Citing such concerns, one Midwest architect designed a ten-story municipal building with all the windows sealed shut. Recognizing the problem, the energy/environmental manager for the city went through and manually changed all those windows to be operable.

Blaming tight buildings gave us charts like the one in Fig. 1, where we were encouraged to compare those minuscule energy savings with the huge personnel losses. The implications were clear: we were trying to save pennies in EE while losing many dollars to lost productivity due to poorer working conditions.

The conclusion seemed to be obvious: there is a direct correlation between EE and IAQ problems.

To prove this hypothesis, however, it is necessary to show that EE buildings have poorer air quality and lower productivity. Or, to state it another way, there is a direct correlation between a tight building and occupant health.

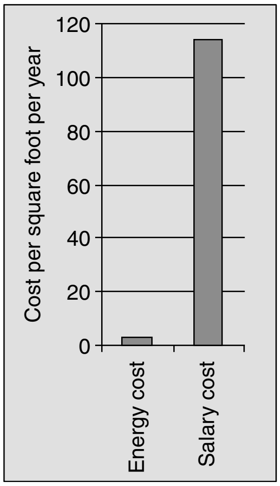


Fig. 1 Cost of energy and salaries in a typical office building.

With a little regression analysis, we ought to be able to build a straight-line relationship: the more energy efficient a building becomes the greater absenteeism and lost productivity become.

As the virtues of tight buildings are weighed, it is easy to forget that those creaky, decrepit old leaky buildings were full of unconditioned, unfiltered, uncontrolled breezes. Drafty buildings were just as apt to cause discomfort as fresh air was.

More than a decade ago, Joseph J. Romm's excellent article "Lean and Clean Management"[4] cited several instances in which EE improved productivity. One example he offered was West Bend Mutual Insurance Company's 40% reduction in energy consumption while it documented a 16% increase in productivity.

VIRTUES OF VENTILATION

Ventilation definitely has its place in an IAQ program. Ventilation can be a good mitigating strategy when the contaminant or its source cannot be determined. Ventilation can also serve as an intermediate step until action can be taken. Further, ventilation may be the best option when source mitigation strategies are simply too costly. Specific applications of ventilation (e.g., localized source control or subslab ventilation to control radon) are valuable control measures. In such instances, more energy may need to be consumed to satisfy IAQ needs.

DISTINGUISHING BETWEEN ENERGY EFFICIENCY AND CONSERVATION

In considering IAQ needs, the distinction between energy conservation and EE becomes critical. By definition, conservation means using less. Further, conservation is still associated with the Emergency Building Temperature Restriction regulations of the 1970s, which led us to equate conservation with deprivation. On the other hand, EE means using the required amount of energy for a healthy, productive workplace or for a process as efficiently as possible.

If we are true to such a definition, it is always possible to have both EE and IAQ.

THE REAL IAQ/EE RELATIONSHIP

There is a surprising relationship between IAQ and EE. First, survey after survey tells us that when utility bills started climbing in the 1970s, the first place where many owners and facility managers found the money to pay those bills was the maintenance budget. This was especially true of institutions on rigid budgets, such as public schools and hospitals. As the utility bills have gone up through the years, those institutions have progressively cut deeper and deeper into maintenance until their deferred maintenance bills have become staggering.

The second relationship between IAQ and EE can also be traced back to energy prices and maintenance. As energy prices climbed, owners bought more sophisticated energy-efficient equipment. Unfortunately, the training of operations and maintenance (O&M) personnel to operate and maintain that equipment did not keep up. Sometimes, the training wasn't offered when the equipment was installed. More often, there was turnover in the O&M personnel, and the new staff did not receive the necessary training.

Keeping these relationships in mind, it's sad to learn that for a long time, we have known that a majority of the IAQ problems found are due to inadequate operations and maintenance. Table 1 offers a review of IAQ problems found by NIOSH, Honeywell's IAQ Diagnostics Group, and the Healthy Buildings Institute (HBI) in the early 1990s. The labels are different, but the commonality of O&M-related problems is very apparent.

THE MUTUAL GOAL OF IAQ AND EE

A careful look at our true goal is needed. Every facility management professional and design professional professes that it is his or her desire to provide owners a facility that has an attractive, healthy, safe, productive environment as cost-effectively as possible. If indeed that is the goal, IAQ and EE are very compatible. They go hand in hand.

Assessing some guidelines of the "1980s" will help bring these two aspects in line. First, let's look at what we call the 80-10-10 rule. Eighty percent of IAQ problems can usually be spotted with an educated eye and a walk-through of a facility. This walk-through may include some very basic measurements (temperature, humidity, CO_2, etc.), but it is not a sophisticated, in-depth investigation. The other 20% of problem facilities require more specialized testing—often exhaustive, expensive

Table 1 Sources of indoor air quality (IAQ) problems

Org.	NIOSH	HONEYWELL	HBI
Bldgs.	529	50	223
Yr.	1987	1989	1989
	Inadequate ventilation (52%)[a]	Operations and maintenance (75%)	Poor ventilation
			No fresh air (35%)
		Energy mgmt.	Inadequate fresh air (64%)
	Inside	Maintenance	Distribution (46%)
	contamination (17%)	Changed loads	
	Outside	Design	Poor filtration
	contamination (11%)	Ventilation/distribution (75%)	Low filter efficiency (57%)
	Microbiological	Filtration (65%)	Poor design (44%)
	contamination (5%)	Accessibility/drainage (60%)	Poor installation (13%)
	Building fabric	Contaminants (60%)	Contaminated systems
	contamination (3%)	Chemical	Excessively dirty
		Thermal	duct work (38%)
		Biological	Condensate trays (63%)
			Humidifiers (16%)

[a]Percentages exceed 100% due to the multifactorial nature of IAQ problems.
Source: From The Fairmont Press (see Ref. 1).

testing, which typically finds only one-half of the remaining problems. To summarize, 80% of IAQ problems are detected through a relatively simple walk-through; 10% are resolved through sophisticated, expensive testing; and nearly 10% remain unresolved.

When considered from the EE perspective, a U.S. Department of Energy study conducted by The Synetics Group (TSG)[5] reported that up to 80% of the savings in an EE program comes from the energy-efficient practices of the O&M staff. What bitter irony! To save money to pay the utility bill, owners cut operations and maintenance. Then they end up with maintenance-related IAQ problems and higher energy bills. So the vicious cycle starts all over again, with more cuts in the maintenance budget.

Fortunately, a positive side to such a vicious cycle can help reverse the situation.

If an IAQ walk-through investigation is paired with a walk-through energy audit, (For more information on audits and auditing, see Hansen and Brown.[2]) it is quite conceivable that the identified future energy savings can pay for the needed IAQ work. One walk-through can identify the IAQ problems and determine ways to finance the mitigation. This approach proves once more how compatible IAQ and EE can be.

EE VS IAQ

In pondering this relationship, it is well to consider what will happen as energy prices continue their upward trend (and they will), if for no other reason than that we need to start calculating the real cost of energy. Whether the increases are due to unrest in the Middle East or increasing demands from China, or whether we start doing a better job of figuring the costs of externalities, prices will trend upward. We are dealing with a finite source and increasing demands. It is a serious miscalculation to assume that fossil fuel prices will have a downward trend.

As the cost of energy goes up, IAQ and EE are apt to be at loggerheads. This does not have to be the case. If IAQ leaders persist in attributing IAQ problems to energy-efficient buildings, as well as relying on more and more outside air for the answer, we lose. For cost-conscious owners, climbing energy costs typically outweigh most IAQ concerns. Only expensive—and often unnecessary—regulations setting outside air requirements can compete against escalating energy prices. The regulated solution may not solve the IAQ problem, but it will definitely increase energy costs—and increase pollution emissions. Sadly, it could all be paid for with money that typically is wasted on energy inefficient operations.[3]

Environmental concerns, higher energy prices, national security issues, and the unnecessary waste of our limited energy resources make increased ventilation a costly answer at the very least. Sustainable development means that we all put our heads together and work for a quality indoor environment, EE, and a quality outdoor environment.

Research has shown that the use of outside air is not always our best option, and that there is nothing wrong with a tight building—provided that a tight building is well designed and well maintained. Tight buildings more readily reveal professional errors. Tight buildings are less forgiving of poor maintenance. A well-designed, well-maintained tight building, however, can provide EE and quality indoor air.

If our ultimate goal is to produce a comfortable, productive indoor environment as cost effectively as possible, EE and IAQ are on the same side. Good managers and effective EE consultants need to have command of both if they are to do their jobs effectively.

REFERENCES

1. Burroughs, H.E.; Hansen, S.J. *Managing Indoor Air Quality*, 3d Ed.; The Fairmont Press: Lilburn, GA, 2005.
2. Hansen, S.J.; Brown, J.W. *Investment Grade Energy Audits: Making Smart Energy Choices*; The Fairmont Press: Lilburn, GA, 2004.
3. Hansen, S.J. *Manual for Intelligent Energy Services*; The Fairmont Press: Lilburn, GA, 2002.
4. Romm, J. Lean and clean management. Energy Environ. Manage. **1995**.
5. The Synetics Group. Evaluation of ICP Grants Program, 1983.

Aircraft Energy Use

K. E. Ohrn
Cypress Digital Ltd., Vancouver, British Columbia, Canada

Abstract
The aviation industry consumes a relatively small amount of the world's fossil fuels. It has a solid record of reducing its consumption and is driven to do so by the large economic impact of fuel costs on the industry. Fuel consumption has been reduced by constant change and improvement in engine and airframe technology, materials, operations planning, aircraft operation, and maintenance practices.

There are incremental gains to be realized in aircraft technology and Air Traffic control technology and procedures. But the predicted rate of industry growth through 2020 will exceed these gains, causing an increase in the industry's overall consumption of fossil fuels. And there do not appear to be any new fuels on the planning horizon.

Several promising areas for breakthrough aircraft technologies have been identified, but all of them are very challenging. Similarly, major gains in Air Traffic Control efficiencies will not be easy to implement.

INTRODUCTION

This article is a broad overview of fuel consumption in the aviation industry. It covers a range of topics, each of which could be expanded considerably. It is intended as an introductory reference for engineers, students, policy-makers, and the general public.

Commercial aviation burns a relatively small 2%–3% of the world's fossil fuels. Military and general aviation accounts for a small and declining proportion of that amount. Since fuel makes up about 12%–17% of an airline's operating costs, the industry has clear economic incentives to reduce consumption. Strong competition in the airline business and its supplier industries has made such progress rapid and effective.

Aviation fuels are comprised mostly of kerosene, which is produced through the distillation of crude oil. Jet turbine fuels account for around 6% of refinery production world-wide. The biggest contribution to the reduction of fuel consumption has been the development of aircraft propulsion from piston-driven propellers to turbofans that use very exotic materials. Air frames have evolved from wood and canvas to aluminium, titanium, and carbon–fiber composites, with significant reduction in weight and an increase in strength.

Once aircraft are in service, operators maintain and operate them effectively in a variety of ways. These include drag reduction programs, flight planning systems, pilot techniques, advanced on-board flight control systems, maintenance, and trend analysis programs.

In the future, incremental gains in fuel efficiency will continue as weight reductions and engine efficiency gains continue, along with the utilization of more sophisticated control systems and manufacturing processes.

Air Traffic Control systems and procedures can be improved, leading to fuel efficiencies through reduced trip distances, and less holding or local maneuvering.

The industry, as a whole, illustrates how far fuel conservation can be taken.

TRENDS IN CONSUMPTION

Total

In 2003,[1] the Transportation sector accounted for 27% of the world's energy consumption, using most of its share as common gasoline. Aircraft account for between 2 and 3% of the fossil fuel burned world-wide, and 6% of petroleum consumption.[2]

The vast majority of aircraft fuel is consumed by the world's 18,000 or so commercial jet aircraft. Fuel consumption by military aircraft is estimated[3] to have dropped from 36% in 1976 to 18% in 1992, and is projected to drop to 7% in 2015 and 3% in 2050.

Most industry associations and observers[4,5] predict a continued growth rate in flight and passenger volumes, averaging at 5% per year for the industry.

Consumption per ATK

The industry has been able to lower its average fuel consumption dramatically in the last 40 years, even as the level of flight activity has soared (Fig. 1).[1,6]

The Available Ton-Kilometer (ATK) is a measure of production capacity for both scheduled and unscheduled passenger and freight aviation. An ATK is defined as the

Keywords: Aircraft; Aviation; Turbine; Fuel; Kerosene.

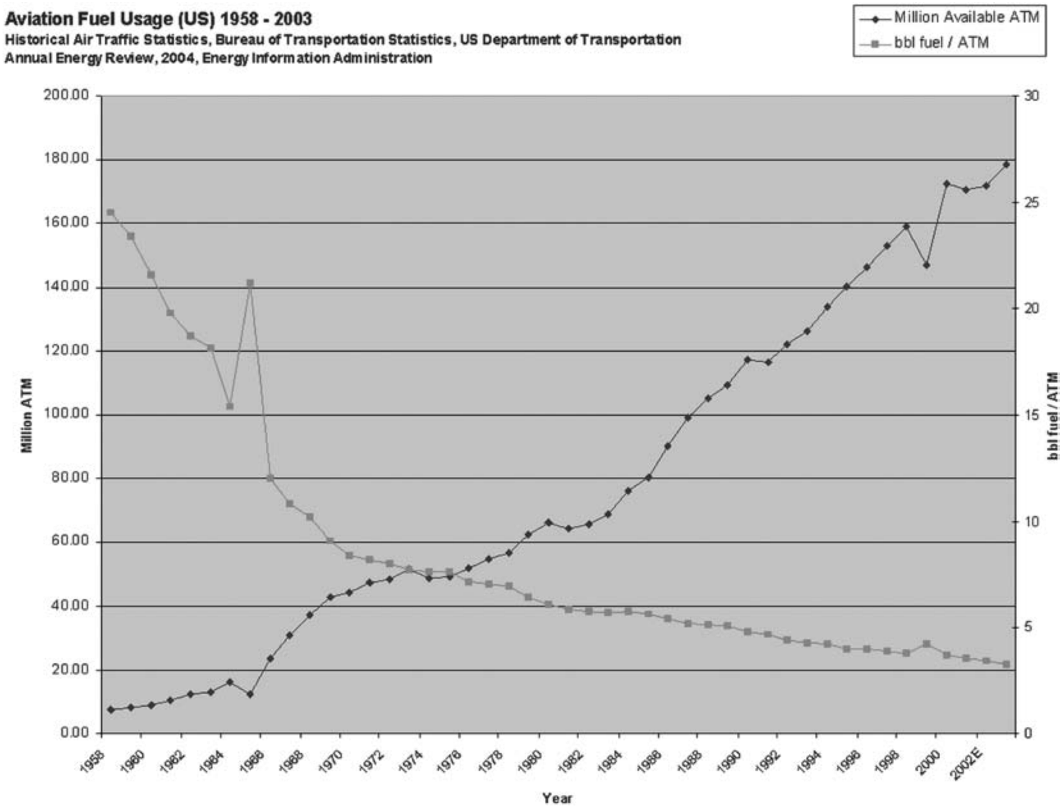


Fig. 1 Aviation fuel consumption per available ton-kilometer (ATK).

number of tons of available aircraft capacity multiplied by the number of kilometers these tons are flown. This measure isolates fuel efficiency discussions of technology and infrastructure from more complex discussions concerning fuel usage per passenger-kilometer, which is more of a market-driven measure. Airlines have several nontechnological paths to pursue in obtaining the most revenue for their fuel dollar. These include keeping aircraft as full as possible, matching aircraft type and schedule to routes and demand, and, thus, spreading fuel costs to cover more passengers and generate greater revenue.

ECONOMIC IMPACT OF ENERGY COSTS ON COMMERCIAL AVIATION

Impetus Towards Conservation

Like many other industries, the energy in aircraft fuel is crucial to airline operations. But fuel is a large percentage (12%–17%) of airline operating costs, usually second only to wages. The percentage varies with the type of carrier and its route structure. Fuel cost varies by as much as 30% between different airports due to transportation costs from refinery to airport, local supplier cost structures, volume discounts, and government tariffs or price support policies.[4]

Given the highly competitive nature of the business, and its high-cost, low-margin characteristics, there is strong reason to pursue fuel efficiency. And the industry has been diligent and successful in its conservation efforts.

In addition to reducing consumption, airlines pursue several strategies to reduce fuel cost. These include fuel tankering, or carrying excess fuel from a low-cost airport to a high-cost one; local supplier negotiations; hedging; and so forth. These strategies extend beyond the scope of this article, but simply reducing consumption still constitutes the best long-term strategy for dealing with fuel costs.

PRODUCTION OF AVIATION FUEL

Crude oil delivered to a refinery is converted into upward of 2000 products,[7] but the most profitable and high-volume products are gasoline, jet fuels, and diesel fuel. Naphtha jet fuels have been used in the military, but phased out in favor of kerosene-based fuels. The major jet fuel types are Jet A, Jet A-1, JP-5, and JP-8.

In terms of overall refinery production, jet fuel accounts for around 6% of output by volume. Only a fairly small number of refineries produce jet fuel.

Sans breakthroughs, kerosene-based fuels seem to be inescapable for the industry.[8]

Jet Fuels

Jet fuel is an output derived from atmospheric distillation, catalytic cracking, and, in some cases, hydro treating sections of the refinery. This depends on the composition of the input crude oil. The final product is a blend of distilled kerosene, which is often upgraded to remove impurities, and heavier hydro and catalytically cracked distillates.[7,9]

The fuel grade (or type) is controlled through strict specifications by American Society for Testing and Materials (ASTM) International, the military, and others.

- *Jet A*—widely available in the U.S.; freeze point −40C; somewhat cheaper and easier to make than other types, helping to ensure wide availability
- *Jet A-1*—widely available outside the U.S.; freeze point −47C. This and Jet A account for the majority of jet fuel usage.[9]
- *Jet B*—blend of gasoline and kerosene (so-called "wide-cut" fuel, which has a range of hydrocarbon compounds in the gasoline and kerosene boiling ranges) as an alternative to Jet A-1 for higher performance in very cold climates
- *JP-4*—military equivalent of Jet B; with additional corrosion inhibition and anti-icing additives (NATO code F-40)
- *JP-5*—military specification (NATO Code F-44)
- *JP-8*—military equivalent to Jet A-1; with additional corrosion inhibition and anti-icing additives (NATO code F-34)

Aviation Gasoline

Gasoline is used in aviation piston engines, and accounts for about 1.1% of the volume of jet fuel, and about 0.2% of the volume of motor fuel[1] in the U.S. ASTM Specification D910 recognizes two octane ratings, 80 and 100, and a low-lead version called 100LL. The product is usually a blend of distillates and alkylates. Normally, the refiner adds tetraethyl lead as required to meet the grade specifications, as well as identifying dyes, which improves safety by allowing different grades of aviation gasoline to be identified by color. The refiner may also add an icing inhibitor, antioxidants, and a few other approved additives, depending on local requirements.

FACTORS IN CONSUMPTION REDUCTION AND CONTROL

Engines (Design History)

According to data produced by Rolls–Royce[10] and International Air Transport Association (IATA),[11] the industry has greatly reduced fuel consumption in the last 50 years. Over 60% of that reduction is due to vast changes in engine technology.

The piston engine which made the 1903 flight of the Wright Brother's first aircraft possible was a water-cooled, four-cylinder, inline design. It weighed around 179 lbs and produced 12 hp.

In 1917, driven by World War I military requirements, the Liberty engine produced 400 hp from an air-cooled V-12 design, and weighed about 790 lbs.

This progress continued through World War II until around 1950, when the Wright R-3350 typified the end-stage of aircraft piston engine development. It produced 3700 hp from 18 cylinders arranged in two radial rows and weighed about 3670 lbs. But the end was in sight due to complexity and limits to the overall power available from a cost-effective piston aircraft engine.

The industry, both military and commercial, had by then turned its attention to the turbine (or "jet") engine. By 1950, Pratt and Whitney had demonstrated the J-57, which produced high thrust with reasonably low fuel consumption. It was used in early jet transports, such as the Boeing 707 and Douglas DC-8. Although very noisy, the engines were acceptably economical to operate and provided the industry much faster aircraft that offered improved travel experiences.

The industry wanted larger engines that generated more power so that much larger aircraft could be built. Engine manufacturers Pratt and Whitney, Rolls–Royce, and General Electric came up with turbofan engines. These engines have cores like traditional turbine engines, but much of their thrust is derived from a large fan located at the front of the engine. This engine can be thought of as an axially-driven, multi-bladed, ducted propeller. This "propeller" provides up to 85% of the engine's thrust. This engine type is characterized by its "bypass ratio," which describes the amount of air coming into the engine that does not go through the central turbine. Typical bypass ratios are from five to one to nine to one. Additionally, these engine designs are much quieter than their predecessors and have much better fuel economy.

A leading engine of this class in 2005 was the General Electric GE90-115B, which weighs 18,300 lbs and produces 120,000 lbs of thrust.

The performance improvements in turbine engine technology are illustrated below. Specific Fuel Consumption (SFC) is a measure of the fuel flow (lbs per hour) of the thrust produced by the engine. The chart shows a sample of representative commercial turbine engines (Fig. 2).

Air Frames (Weight, Composite Materials)

An aircraft in level cruise flight has four forces in balance: lift equals weight and thrust equals drag. Engines provide thrust and consume the aircraft's fuel. For commercial

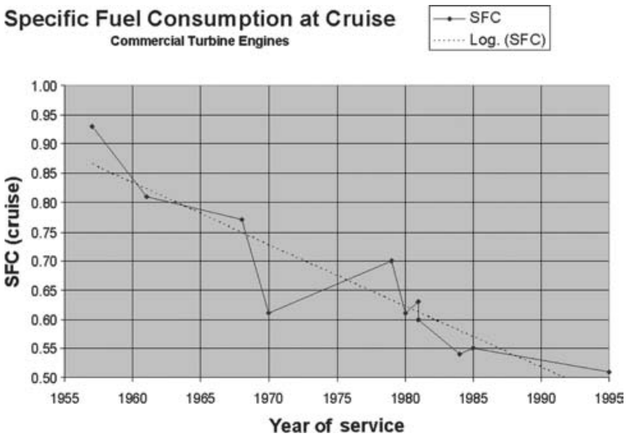


Fig. 2 Specific fuel consumption (SFC) trend chart.

aircraft, drag comes from several sources, but the largest are: induced drag, a by-product of lift, and parasitic drag, which is caused by the air friction and turbulence over the exterior surfaces of the aircraft, as the aircraft moves air out of its way, and by antennae, landing gear, and so on.

Induced drag has a strong relationship to weight: less weight means less lift is required. Induced drag also depends on the design of the wing and its airfoil (wing cross-section) and the angle of attack of the wing. Generally, this drag increases with the square of lift, and the square of aircraft weight.

Any reduction in aircraft weight will directly result in reduced fuel consumption. Reduction in an engine's fuel consumption means that less fuel is required for a given aircraft and payload, hence there is a cumulative effect on its overall efficiency.

Aircraft manufacturers have continuously used a variety of means to reduce aircraft weight. The materials used for aircraft structures have changed from wood and canvas to plywood through various aluminium alloys and, since the 1970s, have included carbon–fiber composites used for simple panels and complex and critical components, such as engine fan blades.

In the Boeing 787 aircraft scheduled for first flight in mid-2007, carbon–fiber resin composites will make up approximately 50% of its structural weight, compared to 12% for the Boeing 777. The planned composite components extend far beyond the usual and into the wings and entire fuselage sections. Aluminium, titanium, and steel constitute the remainder of the structural weight. Airbus Industrie is more conservative, with about 25% composites in the airframe of the A380, which is scheduled to enter service in late 2007.

In the U.S., National Aeronautics and Space Administration (NASA's) Advanced High Temperature Engine Materials Technology Program and the National Integrated High Performance Turbine Engine Technology (IHPTET) Program have investigated and promoted the use of polymer-matrix composites, metal-matrix/intermetallic-matrix, and ceramic-matrix composites for high-temperature parts of aircraft engines. These materials could allow the construction of higher-temperature engines with greater combustion efficiency, all at significantly lower weights. An example is the F136 military engine, which uses a titanium matrix composite in its compressor rotors.

Airlines also work to manage and reduce weight throughout the aircraft and its operation. Excess weight can build up from moisture, dirt, and rubbish in the aircraft, unnecessary supplies, and excess passenger equipment. Boeing[12] estimates that an aircraft will increase in weight by about 0.1%–0.2% per year, leveling off at about 1% in five to ten years. A 1% reduction in weight results in a 0.75 to one percent reduction in trip fuel, depending on the engine type.

Drag Reduction Programs

Drag increases required thrust, so aerodynamic cleanliness is an ongoing challenge.[13,12] Dirt and oil, skin roughness, dents, misaligned fairings, incorrect control rigging, deteriorating seals, mismatched surfaces, and joint gaps (e.g., doors and access panels) all contribute to drag and increased fuel consumption. The most sensitive areas of the aircraft are those where local flow velocities are high and boundary layers are thin: the nose area, the wing leading edges and upper surface, the elevator and rudder leading edges, engine nacelles, and support pylons. If not maintained, a modern transport aircraft can expect a two percent increase in drag within a few years as a result of these factors.

Aircraft Pre-Flight Planning

Before every flight, pilots and operations staff make decisions that affect the overall fuel consumption of each aircraft. Based on knowledge of the aircraft, schedule, payload, and weather, they prepare a load plan and a flight plan. The variables in these plans, and decisions made around them, have a major effect on fuel consumption for the flight.

Center of Gravity (C of G)

Operations dispatchers plan the fuel and cargo load to place the center of gravity within the correct range for safe operation. However, if possible, placing the C of G in the aft portion of this range will result in reduced fuel consumption. This is because when C of G is aft, there is less elevator control surface negative lift required to maintain the correct cruise attitude. This means less lift is required from the wings, resulting in less induced drag. Less negative lift from the tail plane also means less induced drag from this area.

Fuel Quantity

Extra fuel, while comforting to passengers and crew, requires extra fuel burn due to the weight of this extra fuel. A better strategy is to accurately plan the flight to carry the correct amount of fuel and reserves. Elements of this strategy are to:

- Determine the accurate payload and use aircraft weight by tail number if possible.
- Plan the fuel load as required for safety and regulatory requirements, with optimum choice of an alternate airport, careful consideration of the rules that apply to the flight, depending on its origin and destination, and minimal "discretionary" fuel requests.
- If possible, use the re-dispatch technique to minimize contingency fuel requirements.
- Provide accurate, optimized flight planning using the latest origin, destination, and en-route weather information and planning techniques. This involves: choosing a great circle route to reduce distance traveled, if possible; flying pressure patterns and maximizing the use of prevailing wind to reduce enroute flying time; selecting cruise speeds that are, again, an optimum compromise between fuel consumption and schedule performance considerations; and using step climb techniques as required to move to newer altitudes as aircraft weight decreases during the flight.

Pilot Operations Techniques

Pilots can make incremental reductions to fuel consumption through a variety of techniques. They can delay starting the engine until the last minute, after Air Traffic Control (ATC) has issued departure clearances, so that such delays occur at the gate with the engines off. Pilots can minimize the use of the on-board Auxiliary Power Unit, a small turbine that supplies electrical power and compressed air at a higher fuel cost compared to ground power units. Where permitted, the aircraft can also taxi on one engine. Ground operations thrust and braking can then be minimized.

Moreover, pilots can utilize the appropriate flap settings, and retract them as soon as possible to reduce drag. They can follow minimum cost climb profiles whenever possible, but may be thwarted by noise restrictions and ATC congestion problems.

In flight, good control surface trim techniques can save as much as 0.5% in fuel burn by minimizing drag.[12] The appropriate management of air conditioning packs can reduce fuel burn by 0.5%–1.0%. Pilots can use cargo heat and anti-icing judiciously.

There is an optimum point to begin descent into the destination airport. If the plane descends too early, fuel is wasted due to higher consumption while cruising at lower altitudes; if the plane descends too late, the descent speed is too high and energy is wasted. Pilots can delay lowering flaps and landing gear until the last minute: fuel consumption in this high-drag configuration is up to 150% of that in a "cleaner" configuration.

In all of these techniques, safety is the overriding concern. Pilots will always choose a conservative and safe option over a more economical one.

Flight Controls (Autopilot, FMC, W&B)

Modern transport aircraft have significant on-board flight control and management systems that can be used to reduce fuel consumption.

Some Airbus aircraft, for example, have a Fuel Control and Management Computer (FCMC) that can determine the C of G of the aircraft and continuously adjust it toward an optimum position for different flight regimes by pumping fuel to and from an aft-located "trim tank."

Airbus also has a Flight Management Computer (FMC) that can plan step climbs. It also can show the pilots their current optimum altitudes and cruise speeds, in addition to the current actuals, taking into account upper wind forecasts for the flight's planned route. The FMC calculates the optimum top of descent point. When in "managed mode," the FMC uses a "cost index" to account for the carrier's preferences between fuel costs, other direct operating costs, and time savings when calculating cruise altitude and speed.

Control of Engine Maintenance

Boeing[12] recommends several procedures to maintain economical engine operation. These are on-wing washing, which reduces dirt buildup, and bleed air rigging, which compensates for leaks due to system wear. Bleed air is taken from the engine's core and used for a variety of purposes where heated compressed air is needed, such as cabin pressurization and wing de-icing.

Regulatory agencies have mandated significant amounts of on-board data gathering for safety and accident investigation purposes. The industry has found ways to lever this data to provide information about engine health and performance. There are two methods used:

1. Post flight: flight data, gathered manually or electronically, can be loaded into various computer programs after the flight's completion, often on a sampling basis.
2. In-flight: using online data link networks, airlines can downlink in-flight data from the aircraft, among many other types of routine operational reports. ARINC Incorporated of Annapolis, Maryland (GlobalLink) and SITA of Geneva, Switzerland (Aircom) provide Aircraft Communications Addressing and Reporting System (ACARS) services through a world-wide network of satellites,

Very High Frequency (VHF), and High Frequency (HF) ground stations used by airlines and business aircraft operators. Satellite services use four Inmarsat-3 satellites and constitute a global resource for appropriately-equipped aircraft, with the exception of polar regions.

The data are analyzed to determine overall fuel consumption and provide feedback on the success of flight planning and deterioration, if any, of the fuel efficiency of each engine.

This data usually provides the basis for engine trend monitoring, where parameters of interest are compared over time. The onboard computers can also capture short-term "limit exceedance" events, which are gathered on an exception basis.

Airlines, small and large, use in-house software, or software and services provided by many different companies, such as General Electric Aircraft Engines, to perform trend analysis on their engines. This software will predict and characterize trends based on the data provided, including analysis of the combustion efficiency and internal thermodynamics of the hot core sections of the engine. For example, a drop-off in fuel efficiency is probably a sign of wear problems. When certain thresholds of fuel flow, temperature, and so forth are met, the software provides alerts to maintenance staff. In rare cases, an engine may be scheduled for early removal and overhaul. For safety and economic reasons, this is in the operator's best interests. Economic factors include both fuel efficiency and reduced maintenance costs derived as a result of early problem rectification.

THE FUTURE OF GLOBAL AIRCRAFT FUEL CONSUMPTION

Fuel efficiency gains are forecast[3] to be about 2% per year for the foreseeable future. This includes gains from engines, airframes, and operational procedures.

Given a projected airline industry growth rate of about 5% per year, overall industry fuel consumption will continue to rise. If the industry continues to depend upon fossil fuels, it will become more and more expensive and may finally reach a downturn in growth as flights cease to be affordable for tourism and related discretionary travel.

While incremental gains in existing technology are still available, major future gains will depend on breakthrough thinking in airframe design or related technologies.

Incremental Gains

Aircraft designers will continue to reduce aircraft weight through new metals and composites and incremental reduction in the weight of on-board equipment. Active pitch stability features built into fly-by-wire, computer-assisted flight controls (autopilots) could provide a one to three percent reduction in overall fuel efficiency.[8] The continued incorporation of wing-tip devices ("winglets") will reduce induced drag, as will better manufacturing processes, which will smooth exterior surfaces.

Breakthrough Gains

Active systems used to increase laminar flow over the fuselage and wings are very attractive ways to decrease drag, but are fraught with technical challenges.

Fundamentally new designs, such as a blended-wing body, face different challenges, mostly in the realm of passenger acceptance. Similarly, shape-changing wings (morphing-capable) would allow an aircraft to use the most efficient wing size and shape for various flight stages. Coupled with support computers, this could also allow ailerons, rudders, and elevators to be eliminated.

There are potential breakthroughs in materials. Nanotechnology promises to provide materials that are much different and potentially feature orders of magnitude increase in strength-to-weight ratios.

ROLE OF GOVERNMENT

While aircraft technology has been fertile ground for fuel conservation efforts, there are similar efforts underway in other areas. Air Traffic Control is a service that is either regulated by or provided by governments. As such, governments have a large role to play in reducing aircraft fuel consumption.

Clearly, overall trip fuel consumption depends on the length of the flight. If distance traveled and flying time due to holding or local maneuvering can be reduced, optimized, and streamlined, fuel consumption will be reduced. Industry estimates categorize this savings in the six to 12% range over a twenty-year period.[3]

The industry has begun to use Global Positioning Satellites (GPS) to provide optimum point-to-point navigation capabilities. This is in contrast to classic airway navigation, which rarely offers direct or great circle routing, but rather a series of "legs" between fixed-position, ground-based radio navigation stations. When supplemented by a Wide Area Augmentation System (WAAS) and a Local-Area Augmentation System (LAAS), GPS-equipped aircraft can operate in instrument flight conditions for enroute navigation right down to so-called "nonprecision" approaches to the runway. This can reduce distance traveled, fundamentally reducing fuel consumption.

The industry has also continued to move to more advanced ATC systems and procedures. These are intended to streamline airport departure, enroute, and arrival procedures and timing to reduce waste and fuel consumption.

In 1995, the industry began trials of Future Air Navigation System 1 (FANS1) equipment and procedures. This equipment delivers routine ATC information to and from the cockpit via data link, and reduces the use of voice communications, which is a critical bottleneck for air traffic controllers. The industry is moving toward improving arrival and departure sequencing and enroute spacing and increasing flexibility for airline-preferred routing.

Eventually, the industry would like to see a single integrated global air traffic management system to safely optimize the use of scarce airspace (particularly near busy airports).

CONCLUSION

The aviation industry uses a small percentage of the world's energy, but cannot survive in any form without it. The cost of energy in the form of fuel comprises a large percentage of industry operating costs.

In response, the industry has developed considerable expertise and sophistication in monitoring, controlling, and reducing its energy consumption on a per-unit basis. As such, this response serves as an example of how far one can go in pursuit of conservation.

REFERENCES

1. Annual Energy Review 2003. Energy Information Administration, U.S. Department of Energy, DOE/EIA-0384(2003), Washington, DC, U.S., September 2004.
2. Key World Energy Statistics—2004 Edition. International Energy Agency, OECD/IEA—2, rue André-Pascal, 75775 Paris Cedex 16, France or 9, rue de la Fédération, 75739 Paris Cedex 15, France, 2004.
3. Penner, J.E.; Lister, D.H.; Griggs, D.J.; Dokken, D.J.; McFarland, M. Aviation and the Global Atmosphere, Summary for Policymakers, Intergovernmental Panel on Climate Change (IPCC), 1999.
4. Doganis, R. *Flying Off Course, The Economics of International Airlines*, 3rd Ed.; Routledge, Taylor and Francis Group, Oxford, UK, 2002.
5. Technology Update, Aerospace Engineering, March 2003. http://www.sae.org/aeromag/techupdate/03-2003/2-23-2-6.pdf (accessed Sept 20, 2005).
6. Historical Air Traffic Statistics, Annual, Bureau of Transportation Statistics, U.S. Department of Transportation, 400 7th Street, SW, Room 3430, Washington, DC 20590. http://www.bts.gov/programs/airline_information/indicators/airtraffic/annual/1954–1980.html (accessed Sept 20, 2005).
7. Gary, J.H.; Handwerk, G.E. *Petroleum Refining, Technology and Economics*, 3rd Ed.; Marcel Dekker, Inc., 1994.
8. Penner, J.E.; Lister, D.H.; Griggs, D.J.; Dokken, D.J.; McFarland, M. Aviation and the Global Atmosphere, Intergovernmental Panel on Climate Change (IPCC), 1999.
9. Technical Review Aviation Fuels, Chevron Products Company, 2000. http://www.chevron.com/products/prodserv/fuels/bulletin/aviationfuel/pdfs/aviation_fuels.pdf (accessed Sept 20, 2005).
10. Upham, P.; Maughan, J.; Raper, D.; Thomas, C. Towards Sustainable Aviation, Earthscan Publications Ltd, 2003.
11. World Airline Transport Statistics, 2003, IATA.
12. Fuel Conservation, Flight Operations Engineering, Boeing Commercial Airplanes, November 2004. http://www.iata.org/NR/ContentConnector/CS2000/Siteinterface/sites/whatwedo/file/Boeing_Fuel_Cons_Nov04.pdf (accessed Sept 20, 2005).
13. Getting Hand-on Experience with Aerodynamic Deterioration, Airbus Industrie, FAST/Number 21. http://www.content.airbusworld.com/SITES/Customer_services/html/acrobat/fast_21_p15_24.pdf (accessed Sept 20, 2005).

Alternative Energy Technologies: Price Effects

Michael M. Ohadi*
The Petroleum Institute, Abu Dhabi, United Arab Emirates
Jianwei Qi
Department of Mechanical Engineering, University of Maryland, College Park, Maryland, U.S.A.

Abstract
The world is now facing the reality that fossil fuels are a finite resource that will be exhausted someday, that global consumption is outpacing the discovery and exploitation of new reserves, and that the global environment is worsening due to increasing greenhouse gas (GHG) emissions caused by traditional fossil fuels. As a result, there is a renewed push for alternative energy technologies, as well as technologies that can enhance recovery, transportation, and energy utilization and conversion efficiencies. In this entry, a review of current alternative energy technologies and their relevance in various energy sectors will be offered, including their most recent progress and the remaining challenges to overcome. Technology barriers and research/development opportunities for further growth in each category are outlined, and future projected growth is discussed in brief.

ALTERNATIVE ENERGY TECHNOLOGIES

World energy consumption is expected to grow continuously over the next two decades. Much of the growth in new energy demand is expected to come from countries of the developing world, such as Asia. At present, developing countries, comprising more than 75% of the world's population, account for only about one-third of the world's electricity consumption, but this is expected to increase rapidly. Fossil fuels (such as oil, natural gas, and coal) have been the world's primary energy source for several decades due to their competitive low prices. However, with the high world oil prices brought on by the oil price shocks after the OPEC oil embargo of 1973–1974 and the Iranian Revolution of 1979, the use of oil for electricity generation has been slowing since the mid-1970s, and alternative energy sources, such as nuclear power, increased rapidly from the 1970s through the mid-1980s. In addition, given the recent increase in prices of fossil fuels and world compliance with carbon emission reduction policies such as the Kyoto Protocol, nonfossil fuels (including nuclear power and renewable energy sources such as hydroelectricity, geothermal, biomass, solar, and wind power) could become more attractive.

Renewable energy resources have served humans for hundreds of years in the form of water wheels, windmills, and biomass fuels during the industrial revolution. Modern efforts to harness these resources increased sharply after the oil crisis in 1970s, which provided incentive to bring these renewable sources to market to produce electricity for all economic sectors, fuels for transportation, and heat for buildings and industrial processes. Theoretically, renewable energy sources can meet the world's energy demand many times over. After two decades of dramatic technical progress, renewable energy technologies now have the potential to become major contributors to the global energy supply. Some of the technologies are already well established while others require further efforts in research, development, and deployment to become economically competitive in the traditionally fossil fuel-dominated market. However, renewable energy technologies can now be considered major components of local and regional energy systems, as they have become both economically viable and environmentally preferable alternatives to fossil fuels. For each energy demand sector, such as electric power, industrial process and building, and transportation fuels, there are several renewable energy technologies being developed. If any one technology fails to meet the technological and economic goals of its demand sector, at least one other technology will be available for that sector.

Biomass Energy

Biomass refers to green plants or almost any organic product derived from plants. It is actually a form of solar energy that is collected and stored as chemical energy by green plants and then converted to more convenient energy forms (i.e., electric energy and thermal energy) or energy carrier fuels in solid, liquid, and gaseous states. Biomass is the only renewable energy resource that can be converted to liquid fuels like oil. Biomass is used in four

Keywords: Global warming; Greenhouse gases; Renewable energy; Biomass; Solar energy; Wind energy; Hydrogen; Fuel cells.
* On leave from the University of Maryland, College park, where he serves as a Professor of Mechanical Engineering.

main ways: direct combustion, electric power generation, conversion to gas for use as fuel or chemical feedstock, and conversion to liquid fuels. There are abundant biomass resources, including trees and grasses, starch and oil seeds, sawdust, wood waste, agricultural residues, food-processing waste, paper, and municipal solid waste (MSW). Biomass energy commonly refers to both traditional biomass and modern biomass.

Traditional Biomass: traditional biomass is chemical energy directly converted through combustion into thermal energy for heating and cooking. It has been used since humans first discovered fire; it was the first application of renewable energy history. It has taken such forms as fuel wood, animal waste, and crop residue burned in stoves. An estimated 2.4 billion people in developing countries use biomass as their primary fuel for cooking and heating. Traditional biomass provides about 7%–14% of the global primary energy supply and averages 30%–45% of the energy used in developing countries, though some developing countries approach 90%.[1] Today, new biomass stoves and heaters have improved efficiency. About three quads of energy are being provided in the United States today by wood—roughly half the contribution of nuclear power. Municipal solid waste combustion also provides a small amount of process heat. At present, the availability of low-priced wood is the key constraint for its market growth.

Modern Biomass: modern biomass is converted into electricity, transport fuels, or chemicals (e.g., ethanol, methane, and biodiesel) using related process facilities. For example, China and India convert animal and plant wastes into methane for lighting, heating, cooking, and electricity generation, using bacteria to decompose biomass into biogas digesters. Modern biomass accounts for 20% of Brazil's primary energy supply, made possible by significant increases in the past 20 years in the use of ethanol fuels for vehicles and sugarcane waste for power generation. Global annual ethanol production from biomass is estimated at 18 billion liters, 80% of which is in Brazil.[2] Gasification of biomass for methane production may provide a competitive source for the nation's natural gas market. Meanwhile, the conversion of a large portion of MSW and sewage sludge to methane via anaerobic digestion may provide an attractive alternative means of disposing of such wastes.

Biomass Fuels

There are two kinds of biomass-derived liquid fuels for vehicles: ethanol fuel and biodiesel.

Ethanol: ethanol is an alcohol fuel traditionally fermented from corn kernels (corn alcohol). In 2002, 2.13 billion gallons of ethanol were produced in the United States (up 20% from 2001), which was still a small amount compared to U.S. oil imports (less than 2.6%). Ethanol can power specially designed vehicles that run on pure ethanol, or it can be mixed with gasoline or diesel fuel as an additive for use in ordinary vehicles to boost combustion and reduce vehicle emissions. According to the U.S. Environmental Protection Agency, motor vehicle emissions of carbon monoxide can be reduced by 25%–30% with the use of ethanol blended with gasoline. Ethanol–gasoline blends can also reduce ozone levels that contribute to urban smog. In addition, the combustion of ethanol produces 90% less carbon dioxide than gasoline. A blend of 10% ethanol and 90% gasoline has been widely used throughout the nation for many years. Higher level blends of 85 and 95% ethanol are being tested in government fleet vehicles, flexible-fuel passenger vehicles, and urban transit buses. There are already nearly 50,000 such vehicles already in operation, and their use is expected to grow as federal, state, municipal, and private fleet operators seek to comply with the alternative fuel requirements of the Energy Policy Act of 1992 and the Clean Air Act Amendments of 1990.

However, market issues related to ethanol production efficiency, cost competition with gasoline, the commercial viability and costs of specially designed ethanol-only vehicles, the fuel distribution infrastructure, and ratios of ethanol to gasoline in gasohol blending a challenge the attractiveness of ethanol as an alternative fuel. In addition, corn requires high amounts of energy (in the forms of fertilizer, farm equipment fuel, and coal–fire electricity) to grow, harvest, and process. Some research has shown that ethanol consumes more energy than it produces when traditional methods of production are utilized. As a result, renewable energy research has turned its focus toward a new biotech method of producing ethanol—termed bioethanol—from cellulosic biomass, such as agriculture waste products and MSW. Feedstocks include corn husks, rice straws, rice hulls, wood chips, sugarcane, forest thinnings (which also prevent wildfires), waste newspaper, and grasses and trees cultivated as energy crops. Bioethanol requires less energy to produce than ethanol and uses materials that are currently burned or buried. The biological production of ethanol involves hydrolysis of fibrous biomass, using enzymes or acid catalysts, to form soluble sugars, followed by microbial conversion of sugars to ethanol. As a result of technical advances, such as the genetic engineering of specialized enzymes and microbes, the cost of bioethanol production in the lab was decreased from $3.60/gallon in 1980 to about $1.20 in the 1990s.[3] Ultimately, the goal is for bioethanol to become competitive with gasoline in price. Research focuses on producing low-cost enzymes to break down cellulose, improving microorganism performance, producing suitable energy crops, and demonstrating ethanol production from a variety of biomass feedstocks.

Unfortunately, the transition of bioethanol from the laboratory to the highway has been slow. Though biomass is a renewable resource, ethanol is limited by available land. According to a recent Department of Energy (DOE)

Table 1 Four sources required for production of five quads of bioethanol

Source	Portion of ethanol (%)	Note
Cropland	44	Using 10% of current total U.S. cropland, including conservation reserve program acreage
Grassland	19	Using 10% of current grassland
Agricultural waste	25	Using 100% agriculture waste
Waste wood	12	Using 100% waste wood

report, five quads of bioethanol would be needed to provide 45% of the fuel used in gasoline vehicles,[4] which would use current cropland and grassland to produce 63% of bioethanol—as shown in Table 1.

The EIA reports that of the 10.38 MBPD of motor fuel consumed by motor vehicles in the United States in 1999, 0.28 MBPD (2.7%) was comprised of alternative or replacement fuels. More than 90% of this consisted of methyl-tertiary butyl ester (MTBE) (0.2 MBPD) and ethanol (0.06 MBPD) blended with gasoline. Alternative fuels, such as compressed natural gas, methanol, and LPG comprise only 0.02 MBPD.[5] However, MTBE, a petroleum-based oxygenate additive for reformulated gasoline, is toxic and can threaten the safety of community water supplies. When compared with MTBE, ethanol provides extra oxygen, increasing combustion temperatures and efficiency, and lower emissions. Ethanol is expected to replace MTBE as the oxygenate for reformulated gasoline in the near future. As a result, ethanol use increased from 133 trillion Btu in 2001 to 156 trillion Btu in 2002, and surged to 220 trillion Btu in 2003 (see Fig. 1). Production is projected to increase to 278,000 barrels per day in 2025, with about 27% of the growth from conversion of cellulosic biomass (such as wood and agricultural residues) due to rapid improvement in the technology.[6]

Today, more than 60% of Brazil's sugarcane production goes to produce ethanol. Technological advances have continued to improve the economic competitiveness

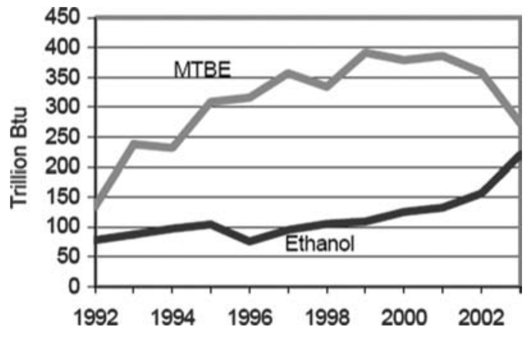


Fig. 1 Ethanol and MTBE consumption in the transportation sector (1992–2003).
Source: From Energy Information Administration (see Ref. 7).

of ethanol and gasohol relative to conventional gasoline, although the price of oil and competitive forces in global automotive technology greatly affects ethanol's prospects. In 2000, over 40% of automobile fuel consumption and 20% of total motor vehicle fuel consumption in Brazil was ethanol, displacing the equivalent of 220,000 barrels of oil per day. Moreover, ethanol is not the only fuel that can be produced from biomass. About 1.2 billion gallons of methanol, currently made from natural gas, are sold in the United States annually, with about 38% of this used in the transportation sector. (The rest is used to make solvents and chemicals). Methanol can also be produced from biomass through thermochemical gasification.

Biodiesel: biodiesel is defined as the mono-alkyl esters of fatty acids processed from any vegetable oil or animal fat. Biodiesel is an alternative fuel for diesel engines that is receiving great attention around the world because it is renewable, and 100% biodiesel eliminates sulfur emissions and reduces particulate matter and other pollutants by 50%. It does increase emissions of one smog-producing pollutant, nitrogen oxide, NO_x, but this can be solved by adjusting engine timing. Biodiesel generally has a lower heating value that is 12% less than No. 2 diesel fuel on a weight basis (16,000 Btu/lb compared with 18,300 Btu/lb). Because biodiesel has a higher density, the lower heating value is only 8% less on a volume basis (118,170 Btu/gallon for biodiesel compared with 129,050 Btu/gallon for No. 2 diesel fuel). Biodiesel can be used in its pure form or in blends with diesel fuel in diesel engines with no modification. In 2000, heavy trucks used 30% as much fuel as light vehicles. Diesel fuel, currently produced from petroleum, is also being produced in limited quantities from soybeans, but research has shown that diesel fuel can also be produced from less costly and more abundant sources, such as the natural oils occurring in algae and pyrolysis of biomass, other vegetable oils (such as corn oil, canola oil, cottonseed oil, mustard oil, and palm oil), animal fats, and waste oils. Total annual production of U.S. fats and oils is about 35.3 billion pounds per year (less than the equivalent of 4.64 billion gallons of biodiesel).[8] The dominant factor in biodiesel production is the feedstock cost, with capital cost contributing only about 7% of the product cost. Sales of biodiesel fuels have exploded thirtyfold since 1999 to 15 million gallons and over 33 million gallons in 2000, per the EIA. The United States has the capacity to produce much more biodiesel, and its output is projected to grow by 1.8% per year.[9]

United States farms and fields have yielded some homegrown energy choices, like biodiesel and ethanol, but it is still hard for them to compete with fossil fuel, with its relatively low prices and robust infrastructure, in the market. Currently, a subsidy is offered by the Department of Agriculture's Commodity Credit Corporation for the promotion and production of biodiesel. However, biodiesel is as much as twice the cost of petroleum. It is obvious that biodiesel could not completely replace petroleum-based diesel fuel in the near future. Even with the unrealistic scenario that all of the vegetable oil and animal fat currently produced were used to produce biodiesel, only about 15% of the current demand for on-highway diesel fuel could be replaced.

Solar Energy

The sun's energy is the primary source for most energy forms found on the earth. Nature's energy resources are confined to two categories: earth-stored fossil residues and nuclear isotopes that are limited by the finite amounts that exist on the earth, and the radiation flux of solar energy that is clean, abundant, and renewable. Although solar energy holds tremendous potential to benefit the world by diversifying its energy supply, reducing dependence on fossil fuels, improving the quality of the global environment, and stimulating the economy by creating jobs in the manufacture and installation of solar energy systems, solar energy's economic utility is limited by the finite rate at which the sun's energy can be captured, concentrated, stored, and/or converted for use in the highest value energy forms, and by the land areas that societies can dedicate to harness it. The amount of solar energy received across U.S. latitudes is approximately 22 quads per year per 4000 km^2 (about a million acres) on average.[10] Thus, about 40–80 thousand km^2 of land—roughly two to four times the size of Massachusetts—could supply about 20 quads, or 20%–25%, of today's total U.S. energy requirements (currently, PV solar cells convert 10%–20% of incident radiation directly to electricity).

Solar Energy Prospects

According to the outline of the U.S. DOE energy technologies program,[11] solar energy will increase the world's energy supply and enhance the reliability of the global energy infrastructure, thus creating a more stable environment for economic growth. The distributed, modular characteristics of solar energy offer tremendous flexibility for both grid-connected and off-grid electricity applications. Distributed energy technologies are expected to supply an increasing share of the electricity market to improve power quality and reliability problems, such as power outages and disturbances. With improved technology supported by the U.S. DOE, the cost of solar energy has dropped substantially in the past decade and continues to decline. The projected costs (shown as dashed lines in Fig. 2) are based on continuing the proposed budget support for the DOE Solar Program.

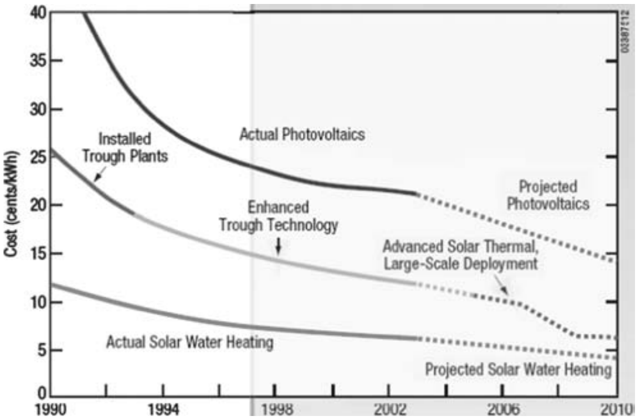


Fig. 2 The cost roadmap of solar energy technology.
Source: From U.S. Department of Energy (see Ref. 11).

The long-term cost goals are even more ambitious. For example, the goal for photovoltaics, which will become an economically competitive alternative to traditional fossil fuel energy, is $0.06 per kilowatt-hour (kWh) in 2020.

Wind Energy

Wind is a form of solar energy. Winds are caused by the uneven heating of the atmosphere by the sun, the irregularities of the earth's surface, and rotation of the earth. Wind flow patterns are modified by the earth's terrain, bodies of water, and vegetation. Wind energy or wind power describes the process by which the wind is used to generate mechanical power or electricity. This mechanical power can be used for specific tasks (such as grinding grain or pumping water) or fed to a generator that can convert this mechanical power into electricity. Since early recorded history, people have been harnessing the energy of the wind. Wind propelled boats along the Nile River as early as 5000 B.C.; by 200 B.C., simple windmills in China were pumping water.[2] Commonly called wind turbines, machines that convert the kinetic energy in the wind into electricity through a generator appeared in Denmark as early as 1890. In the 1940s, the largest wind turbine of the time began operating on a Vermont hilltop known as Grandpa's Knob, which rated at 1.25 MW in winds of about 30 mph and fed electric power to the local utility network for several months during World War II.[12]

The popularity of using wind energy has always fluctuated with the price of fossil fuels. When fuel prices fell after World War II, interest in wind turbines declined. But when the price of oil skyrocketed in the 1970s, so did worldwide interest in wind turbine generators. The rapid progress in wind turbine technology has refined old ideas and introduced new ways of converting wind energy into useful power. Many of these approaches have been demonstrated in "wind farms" or "wind power plants" (groups of turbines), which feed electricity into the utility grid in the United States and Europe. Since the 1970s, wind energy has expanded its role in electricity generation. The worldwide installed capacity of grid-connected wind power has now exceeded 40 GW, corresponding to an investment of approximately $40 billion.[12] A demand for clean, diverse sources of electricity and state and federal incentives to stimulate the market have contributed to wind energy's growth in the United States. Wind energy installations in the United States increased during the past decade from about 1800 MW in 1990 to more than 6000 MW at the end of 2003, enough to power almost three million average homes. The average U.S. wind energy growth rate for the past five years was 24%. This growth can be attributed to a greatly reduced cost of production, from 80 cents (current dollars) per kWh in 1980 to cents per kWh in 2002. The global wind energy installed capacity has increased exponentially over a 25-year period, and in the process, the cost of energy (COE) from wind power plants has been reduced by an order of magnitude, becoming very close in cost to power from modern combined-cycle power plants in some locations. According to the American Wind Energy Association, as much as 13,500 additional megawatts of wind capacity may be installed worldwide in the next decade.[23] Wind energy is the world's fastest growing energy source and will power industry, businesses, and homes with clean, renewable electricity for many years to come.

Wind Energy Technology Development

The considerable potential of wind energy was not tapped before the 1980s because the wind turbine technology was not competitive with most central fossil fuel-fired generating stations. But over the past two decades, the rapid progress in wind turbine technologies has led to more cost-effective wind turbines that are more efficient in producing electricity. The progress was mainly motivated by the oil embargoes and fuel price escalations of the 1970s and more recently by environmental concerns. The goal is to develop cost-effective, low wind-speed turbines for Class sites (13-mph average annual wind speed) that can produce electricity onshore for $0.03/kWh and offshore for $0.05/kWh by the end of 2012. This will open up 20 times more U.S. land for wind energy development, and because many of these sites tend to be closer to urban load centers, the problem of transmission line expansion will be greatly simplified. But the current turbine designs are not well suited to low wind-speed sites and have only limited potential to achieve lower energy costs. If such technology can be successfully developed, the wind resources across the Great Plain states could potentially generate more electricity than is currently consumed by the entire nation.[13] Although wind power plants have relatively little impact on the environment compared to other conventional power plants, there is some concern over the noise produced by the rotor blades, aesthetic (visual) impacts, and the danger to birds, which have been killed by flying into the rotors. Most of these problems have been resolved or greatly reduced through technological advances or by properly siting wind plants.

Future technology improvements for low-speed wind technology must address three principal areas:

- Turbine rotor diameters must be larger to harvest the lower-energy winds from a larger inflow area without increasing the cost of the rotor.
- Towers must be taller to take advantage of the increased wind speed at greater heights.
- Generation equipment and power electronics must be more efficient to accommodate sustained light wind operation at lower power levels without increasing electrical system costs.

Hydrogen and Fuel Cells

Hydrogen is the simplest element; an atom consists of only one proton and one electron. It is also the most plentiful element in the universe. Although in many ways hydrogen is an attractive replacement for fossil fuels, it does not occur in nature as the fuel H_2. Rather, it occurs in chemical compounds, such as water or hydrocarbons that must be chemically transformed to yield H_2. Hydrogen is found in water and hydrocarbon organic compounds that make up many of our fuels, such as gasoline, coal, natural gas, methanol, and propane. Although President Bush has called hydrogen a "pollution free" technology, extracting hydrogen from its most common source, water, requires electricity that may come from fossil fuels, such as coal or nuclear energy. Hydrogen, like electricity, is a carrier of energy, and like electricity, it must be produced from a natural resource. Nevertheless, it promises substantial contributions to global energy supplies and minimal environmental impact in the long term.

Hydrogen

Hydrogen is the simplest chemical fuel (essentially a hydrocarbon without the carbon) that makes a highly efficient, clean-burning energy carrier and a secondary form of energy that has to be produced like electricity. When hydrogen is used to power a special battery called a fuel cell, its only waste product is water. Hydrogen-powered fuel cells and engines could become as common as the gasoline and diesel engines of the late 20th century and could power cars, trucks, buses, and other vehicles, as well as homes, offices, and factories. Hydrogen has the potential to fuel transportation vehicles with zero emissions, provide process heat for industrial processes, supply domestic heat through cogeneration, help produce electricity for centralized or distributed power systems, and provide a storage medium for electricity from renewable energy sources. Some envision an entire economy based on hydrogen in the future.[14] At present, most of the world's hydrogen is produced from natural gas by a process called steam reforming. However, producing hydrogen from fossil fuels would not be an advancement because steam reforming does not reduce the use of fossil fuels, but rather shifts them from end use to an earlier production step; in other words, steam reforming would still release carbon to the environment in the form of CO_2. Thus, to achieve the benefits of the hydrogen economy, the hydrogen must be produced more cost effectively from nonfossil resources, such as water, using a renewable energy source like wind or solar. Although the potential benefits of a hydrogen economy are significant, many barriers to commercialization—technical challenges, and otherwise—must be overcome before hydrogen can offer a competitive alternative for consumers.

Commercial Barriers. The commercial barriers to the widespread use of hydrogen are the high cost of hydrogen production, low availability of hydrogen production systems, the challenge of providing safe production and delivery systems (i.e., economical storage and transportation technologies), and public acceptance.

- *Hydrogen Storage*. Hydrogen has a low energy density in terms of volume, making it difficult to store amounts adequate for most applications in a reasonably-sized space. This is a particular problem for hydrogen-powered fuel cell vehicles, which must store hydrogen in compact tanks. Hydrogen is currently stored in tanks as a compressed gas or cryogenic liquid. The tanks can be transported by truck or the compressed gas can be sent across distances of less than 50 miles by pipeline. Other options are to store hydrogen in a cryogenic liquid state or solid state. Technologies that store hydrogen in a solid state are inherently safer and have the potential to be more efficient than gas or liquid storage. These are particularly important for vehicles with on-board storage of hydrogen. High-pressure storage tanks are currently being developed, and research is being conducted into the use of solid-state storage technologies, such as metal hydrides, which involve chemically reacting the hydrogen with a metal; carbon nanotubes, which take advantage of the gas-on-solids adsorption of hydrogen and retain high concentrations of hydrogen; and glass microspheres, which rely on changes in glass permeability with temperature to fill the microspheres with hydrogen and trap it there. However, the statistical cost, durability, fast-fill, discharge performance, and structural integrity data of hydrogen storage systems must be improved before proceeding with commercialization.
- *Safety, Codes, and Standards*. Hydrogen, like gasoline or any other fuel, has safety risks and must be handled with due caution. Unlike the handling of gasoline, handling hydrogen will be new to most consumers. Therefore, developers must optimize new fuel storage and delivery systems for safe everyday use, and consumers must become familiar with hydrogen's properties and risks. Codes and standards are needed to ensure safety as well as to commercialize hydrogen as a fuel.
- *Public Acceptance*. Finally, public acceptance of hydrogen depends not only on its practical and commercial appeal, but also on its record of safety in widespread use. Because a hydrogen economy would be a revolutionary change from the world we know today, educating the general public, training personnel in the handling and maintenance of hydrogen system components, adopting codes and standards, and developing certified procedures and training manuals for fuel cells and safety standards will foster hydrogen's acceptance as a fuel.

Technology Roadmap (Present-2030). Technical challenges for hydrogen commercialization include cost-effective, energy-efficient production technologies and safe, economical storage and transportation technologies. The U.S. DOE has provided a national version of America's transition to a hydrogen economy to 2030 or beyond.[15] This technology roadmap consists of three steps. The first step toward a clean energy future will focus on technology development and initial market penetration to build on well-known commercial processes for producing, storing, transporting, and using hydrogen. In the mid-term, as hydrogen use increases and hydrogen markets grow, the expansion of the market and infrastructure investment will make the cost of hydrogen and fuel cell economically competitive with traditional fossil fuels. For the long term, when the market and infrastructure are more fully developed, wider uses of more cost-effective advanced technologies will be an important step toward a hydrogen economy.[15]

Today, large centralized steam methane reformers are used to produce hydrogen for chemical industries. This will continue to be the likely choice for meeting increased hydrogen demand in the near term. Electrolyzers are also used to produce the high-purity hydrogen needed for electronics manufacturing and other specialty uses. Compressed hydrogen tanks are available today, although the low energy density of hydrogen means large tanks are needed. As a liquid, hydrogen's energy density is substantially improved, but boil-off losses are a concern. Today, hydrogen is transported by pipeline or over the road in cylinders, tube trailers, and cryogenic tankers.

A small amount is shipped by rail car or barge. Hydrogen has also long been used in the space program as a propellant for the space shuttle and in the on-board fuel cells that provide the shuttle's electric power. New combustion equipment is being designed specifically for hydrogen in turbines and engines, and vehicles with hydrogen internal combustion engines have been demonstrated. Also being tested is the combustion of hydrogen–natural gas blends to improve the yield of natural gas reforming in an effort to lower cost and raise efficiency. Fuel cells are in various stages of development for transportation, stationary, and portable applications. Incremental advances of current technologies provide a low-risk commercial entry into the hydrogen economy.

Fuel Cells

The widespread use of hydrogen as an energy source in the world could help address concerns about energy security, global climate change, and air quality. Fuel cells are an important enabling technology for a future hydrogen economy and have the potential to revolutionize power generation, offering cleaner, more efficient alternatives to the combustion of gasoline and other fossil fuels. Fuel cells promise to be a safe and effective way to use hydrogen for both vehicles and electricity generation. Although these applications would ideally run off pure hydrogen, in the near term they are likely to be fueled with natural gas, methanol, or even gasoline. If the fuel cell is to become the modern steam engine, basic research must provide breakthroughs in understanding, materials, and design to make a hydrogen-based energy system a vibrant and competitive force. Fuel cell technology is not a new invention. Actually, fuel cell development predated the internal combustion engine, but lacked a commercial venue until NASA decided to incorporate fuel cells in spacecrafts during the 1960s. Phosphoric acid fuel cells are already commercially available and can generate electricity in 200-kW capacities selling for $3/W, using natural gas as the source of hydrogen; molten carbonate has also been demonstrated at large (2-MW) capacities.

A fuel cell works like a battery but does not run down or need recharging. Fuel cells convert hydrogen—hydrogen gas or hydrogen reformed within the fuel cell from natural gas, alcohol fuels, or some other source—directly into electrical energy with no combustion. They will produce electricity and heat as long as fuel (hydrogen) is supplied. A fuel cell consists of two electrodes, a negative electrode (or anode) and a positive electrode (or cathode), sandwiched around an electrolyte. Hydrogen is fed to the anode, and oxygen is fed to the cathode. Activated by a catalyst, hydrogen atoms separate into protons and electrons, which take different paths to the cathode. The electrons go through an external circuit, creating a flow of electricity. The protons migrate through the electrolyte to the cathode, where they reunite with oxygen and the electrons to produce water and heat.

Fuel cell technologies are significantly more energy efficient than combustion-based power generation technologies. A conventional combustion-based power plant typically generates electricity at efficiencies of 33%–35%, while fuel cell plants can generate electricity at efficiencies of up to 60%. When fuel cells are used to generate electricity and heat (cogeneration) they can reach efficiencies of up to 85%. Internal-combustion engines in today's automobiles convert less than 30% of the energy in gasoline into power that moves the vehicle. Vehicles using electric motors powered by hydrogen fuel cells will be much more energy efficient, utilizing 40%–60% of the fuel's energy.[28]

Technology Challenges. Although NASA has used hydrogen fuel cells for space missions since the 1960s, terrestrial applications are still in their infancy. The lack of an economical process for hydrogen production and suitable storage methods are two of the greatest obstacles to commercialization, especially in the transportation sector. Research goals include developing technologies to produce hydrogen from sunlight and water and biomass; developing low-cost and low-weight hydrogen storage technologies for both stationary and vehicle-based applications, such as carbon nanotubes and metal hydrides; and developing codes and standards to enable the widespread use of hydrogen technologies. Technological development is addressing the following key challenges in the commercialization of fuel cell and hydrogen infrastructure technologies[28]:

- *Cost*. Cost is the greatest challenge to fuel cell development and adaptation, and it is a factor in almost all other fuel cell challenges, as well. Materials and manufacturing costs are high for fuel cell components (i.e., catalysts, membranes, bipolar plates, and gas diffusion layers). Statistical data for fuel cell vehicles that are operated under controlled, real-world conditions are very limited and often proprietary. For example, some fuel cell designs require expensive, precious-metal catalysts, and others require costly materials that are resistant to extremely high temperatures. Currently, the costs for automotive internal combustion engine power plants are about $25–$35/kW. The targeted cost for a fuel cell system to be competitive in transportation applications is around $30/kW, and the acceptable price point for stationary power systems is considerably higher (i.e., $400–$750/kW for widespread commercialization and as much as $1000/kW for initial applications).
- *Durability and Reliability*. All fuel cells are prone, in varying degrees, to catalyst poisoning, which decreases fuel cell performance and longevity. The durability and reliability of fuel cell systems operating over automotive drive cycles has not been established. Vehicle

fuel cell power systems will be required to achieve the same level of durability and reliability of current engines over the full range of vehicle operating conditions at 40°C–80°C (i.e., 5000 h lifespan or about 150,000 miles equivalent). Stationary fuel cells must achieve greater than 40,000 h of reliable operation at −35 to 40°C to meet market requirements. Fuel cell component degradation and failure mechanisms are not well understood. The cycle life of hydride storage systems also needs to be evaluated in real-world circumstances.

- *System Size*. The volume and weight of current fuel cell systems are too high to meet the packaging requirements for transportation or stationary applications. System volume minimization and weight reduction will focus not only on the fuel cell stack, but also on the ancillary components and major subsystems making up the balance of power system (e.g., fuel processor, compressor/expander, and sensors).
- *Air, Water, and Thermal Management*. Fuel cell performance and efficiency must meet or exceed that of competing technologies in order to be commercially accepted. Today's compressor technologies are not suitable for fuel cell applications that need low power consumption and less packaging volume. Vehicle fuel cell systems must start rapidly from any ambient condition with minimal fuel consumption. Cost-effective thermal and water management technologies are needed, including heat recovery and water utilization, cooling, and humidification. The low operating temperature of proton exchange membrane (PEM) fuel cells results in a relatively small difference between the operating and ambient temperatures, which need advanced technologies to allow high operating temperatures and to improve combined heat and power system performance.

Nuclear Energy

Nuclear technology uses the energy released by splitting the atoms of certain elements. It was first developed in the 1940s, and during the Second World War, research initially focused on producing bombs by splitting the atoms of either uranium or plutonium. Only in the 1950s did attention turn to peaceful applications of nuclear fission, notably power generation. Today, the world produces as much electricity from nuclear energy as it did from all sources combined in 1960. Civil nuclear power, with some 440 commercial nuclear power reactors in 30 countries, can now exceed 12,000 reactor years of experience, and nuclear power supplies 16% of global needs with a total installed capacity of over 360,000 MWe.[18] This is more than three times the total generating capacity of France or Germany from all sources. The economics of nuclear power may be more favorable in countries where other energy fuels (mostly imported) are relatively expensive. In 2002, nineteen countries depended on nuclear power for at least 20% of their electricity generation (Fig. 3), while three quarters of both France's and Lithuania's power are derived from nuclear energy.

However, accidents at Three Mile Island in the United States in 1979 and at Chernobyl in the Soviet Union in 1986 pushed public opinion and national energy policies away from nuclear power as a source of electricity. But, after nearly two decades of the antinuclear tide, nuclear energy today is at a turning point. Nuclear power plant

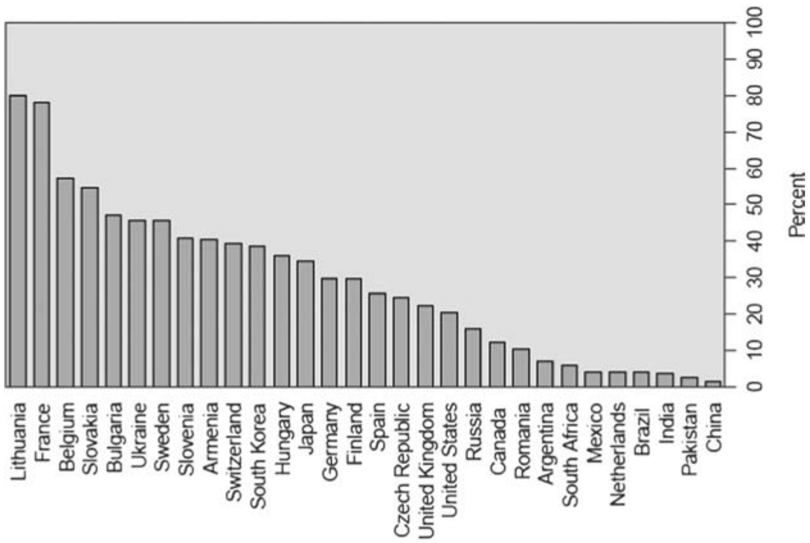


Fig. 3 Nuclear shares of national electricity generation, 2002.
Source: From IAEA, International Atomic Energy Agency (see Ref. 16).

performance has shown a steady improvement over the past 10–15 years: one quarter of the world's reactors have load factors of more than 90% and almost two-thirds do better than 75%, compared to about a quarter of them in 1990.

Nuclear power is the only mature, noncarbon electricity generation technology that can significantly contribute to the long-term, globally sustainable energy mix. Besides providing electricity, nuclear energy contributes to a number of policy goals, including achieving energy independence, keeping the air clean, and reducing carbon emissions. European countries have begun construction of a nuclear reactor, and six more are likely to be constructed in the next decade. The U.S. Nuclear Power 2010 program aims to build new nuclear power plants in the United States by the end of the decade, and expects that the advanced reactor designs will produce electricity in the range of $1000–$1200 per kW of electricity. The highest growth in nuclear generation is expected in the developing world, where consumption of electricity from nuclear power is projected to increase by 4.1% per year between 2001 and 2025. Developing Asia, in particular, is expected to see the largest increase in nuclear generating capacity, accounting for 95% of the total increase in nuclear power capacity for the developing world. Of the 44 gigawatts of additional installed nuclear generating capacity projected for developing Asia, 19 gigawatts are projected for China, 15 gigawatts are projected for South Korea, and 6 gigawatts are projected for India.[18]

Nuclear energy is, in many places, competitive with fossil fuel for electricity generation, despite relatively high capital costs and the need to internalize all waste disposal and decommissioning costs. If the social, health, and environmental costs of fossil fuels are also taken into account, nuclear energy is superior. A 2004 report from the University of Chicago, funded by the U.S. DOE, compares the power costs of future nuclear, coal, and gas-fired power generation in the United States. Various nuclear options are covered, and for ABWR or AP1000, they range from 4.3 to 5.0 c/kWh on the basis of overnight capital costs of $1200–$1500/kW, a 60-year plant life, a 5-year construction period, and 90% capacity. Coal yields 3.5–4.1 c/kWh and gas (CCGT) 3.5–4.5 c/kWh, depending greatly on fuel price.[17] When considering a minimal carbon control cost impact of 1.5 c/kWh for coal and 1.0 c/kWh for gas superimposed on the above figures, nuclear is even more competitive.

Overview of Nuclear Power Technology

The principles for using nuclear power to produce electricity are the same for most types of reactors. The energy released from continuous fission of the fuel atoms is harnessed as heat in either gas or water and it is used to produce steam. The steam is used to drive the turbines, which produce electricity (as in most fossil fuel plants). In most naval reactors, steam drives a turbine directly for propulsion.

There are several components common to most types of reactors:

- *Fuel*. Pellets of uranium oxide (UO_2) are usually arranged in tubes to form fuel rods. The rods are arranged into fuel assemblies in the reactor core.
- *Moderator*. This slows down the neutrons released from fission so that they cause more fission. The moderator is usually water, but may be heavy water or graphite.
- *Control Rods*. These are made from neutron-absorbing material, such as cadmium, hafnium, or boron, and are inserted into or withdrawn from the core to control the rate of reaction, or to halt it.
- *Coolant*. A coolant, such as a liquid or gas, circulates through the core so as to transfer heat from it. In light water reactors, the moderator also functions as coolant.
- *Pressure Vessel/Pressure Tubes*. This is a robust steel vessel containing the reactor core and moderator/coolant, but it may also be a series of tubes holding the fuel and conveying the coolant through the moderator.
- *Steam Generator*. This is the part of the cooling system where the heat from the reactor is used to make steam for the turbine.
- *Containment*. The packaging structure around the reactor core is designed to protect it from outside intrusion and to protect those outside from the effects of radiation in case of any malfunction inside. It is typically a meter-thick concrete and steel structure.

Types of Nuclear Reactors

Most nuclear electricity is generated using just two kinds of reactors that were developed in the 1950s and have been improved since. In the United States, Westinghouse designed the first fully commercial pressurized water reactor (PWR) of 250 MWe, Yankee Rowe, which started up in 1960 and operated until 1992. Meanwhile, Argonne National Laboratory developed the first boiling water reactor (BWR). The first of this type, the Dresden-1 of 250 MWe, was designed by General Electric and started earlier in 1960. A prototype BWR, Vallecitos, ran from 1957 to 1963. By the end of the 1960s, orders were being placed for PWR and BWR reactor units of more than 1000 MWe.[19]

Pressurized Water Reactors utilize pressurized water as a moderator and coolant. The fuel, ceramic uranium dioxide, is typically encased in long zirconium alloy tubes. The uranium-235 is enriched from its original 0.7% abundance to 3.5%–5.0%.

Boil Water Reactors are similar to PWRs, except that the coolant water is allowed to boil, and steam passes from

Table 2 Overview of commercial nuclear power reactors

Reactor type	Main countries	Number	GWe	Fuel	Coolant	Moderator
Pressurized water reactor (PWR)	United states, France, Japan, Russia	268	249	Enriched UO_2	Water	Water
Boiling water reactor (BWR)	United states, Japan, Sweden	94	85	Enriched UO_2	Water	Water
Gas-cooled reactor (magnox &AGR)	UK	23	12	Natural U (metal), enriched UO_2	CO_2	Graphite
Pressurized heavy water reactor "CANDU" (PHWR)	Canada	40	22	Natural UO_2	Heavy water	Heavy water
Light water graphite reactor (RBMK)	Russia	12	12	Enriched UO_2	Water	Graphite
Fast neutron reactor (FBR)	Japan, France, Russia	4	1	PuO_2 and UO_2	Liquid sodium	None
	TOTAL	441	381			

Source: From Nuclear Engineering (see Ref. 19).

the top of the reactor directly to the turbine. Currently, more than 90 of these are operating throughout the world.

Advanced Gas-Cooled Reactors (AGRs) are the second generation of British gas-cooled reactors, using graphite moderators and carbon dioxide as coolant. The fuel is uranium oxide pellets, enriched to 2.5%–3.5%, in stainless steel tubes. The carbon dioxide circulates through the core, reaching 650°C, and then past steam generator tubes outside the core, but still inside the concrete and steel pressure vessel. Control rods penetrate the moderator, and a secondary shutdown system involves injecting nitrogen into the coolant.

Pressurized Heavy Water Reactors (PHWRs) are a Canadian reactor development headed down a quite different track, using natural uranium fuel and heavy water as a moderator and coolant. The first unit started up in 1962.

Fast Neutron Reactors (FBRs) (there is only one in commercial service) do not have a moderator and utilize fast neutrons, generating power from plutonium, while simultaneously making more of it from the U-238 isotope in or around the fuel. While these reactors get more than 60 times as much energy from the original uranium compared to normal reactors, they are expensive to build and must contend with resource scarcity before coming into their own.

Several different types of reactors in current commercial nuclear power plants are summarized in Table 2.

Several generations of reactors can be commonly distinguished. Generation I reactors were developed in the 1950s–1960s, and relatively few are still running today. They mostly used natural uranium fuel and graphite moderators. Generation II reactors developed in the 1970s–1980s are typified by the present U.S. fleet and most are in operation elsewhere. They typically use enriched uranium fuel and are mostly cooled and moderated by water. Pressurized water reactors and BWRs are known as light-water reactors. Around the world, with few exceptions, other countries have chosen light-water designs for their nuclear power programs, so that, today, 65% of the world capacity is PWR and 23% BWR.

Generation III reactors are advanced reactor designs developed in the 1990s. With enhanced safety, these reactors are more economical to build, operate, and maintain than the previous generation. The first Generation III system, a General Electric-designed advanced BWR, started operating in 1996 at the Kashiwazaki-Kariwa Nuclear Power Station in Japan. More than a dozen Generation III advanced reactor designs are in various stages of development. Some have evolved from the PWR, BWR, and CANDU designs above; some are more radical departures. The best-known radical new design is the Pebble Bed Modular Reactor (PBMR), which uses high-temperature helium to cool the reactor and drive the turbine directly. One of the limitations of current light-water reactor technology is that the thermal efficiency that can be achieved is limited to the achieved maximum temperature of 350°C. The PBMR is designed to achieve at least 900°C, which will give a thermal efficiency of up to 44%. This translates into roughly one-third more output than a conventional PWR. The first PBMR is currently planned for commercial operation in South Africa by around 2010.

Future Nuclear Power Technology

Generation I–III reactors recycle plutonium (and possibly uranium), while the future of nuclear reactors, known as Generation IV systems, have revolutionary reactor and fuel cycle systems.[29] Most will have closed fuel cycles that burn the long-lived actinides that form part of the spent fuel so that fission products are the only high-level waste. Many will also be fast neutron reactors. Six new designs, including three PWRs, three BWRs, and two high temperature gas-cooled reactors (HTRs), were identified for further study by the Generation IV International Forum (GIF), which was initiated in 2000 by a group of nine countries: Argentina, Brazil, Canada, France, Japan, South Africa, South Korea, the U.K., and the U.S. Switzerland became a member of the forum in February 2002, and the European Atomic Energy Community joined in July 2003. GIF is an international initiative whose goal is to develop nuclear energy systems that can supply future global needs for electricity, hydrogen, and other products. The aim is to deploy these systems no later than 2030 to provide competitively priced and reliable energy products, while satisfactorily addressing nuclear safety, waste, proliferation, and physical protection concerns. All six revolutionary nuclear reactor technology concepts identified for development by GIF operate at higher temperatures than the Generation II and III reactors currently in operation. The new systems range from a supercritical water-cooled reactor (SCWR), which operates at 510°C–550°C, to a helium-cooled very high-temperature gas reactor (VHTR), which has an operating temperature of 1000°C. Three of the six Generation IV concepts are fast reactor systems that are cooled either by helium gas, lead, or sodium. All use depleted uranium as a fuel. The main technological challenges are addressed as follows:

- Generation IV technology must address the high-level waste from fission reactions. This waste includes heavy nuclides–actinides such as neptunium, americium, and curium—that remain highly radioactive for tens of thousands of years. The helium-, lead-, and sodium-cooled fast reactors are designed to have closed fuel cycles in which the actinides are separated from the spent fuel and returned to the fission reactors. Supercritical water-cooled reactors are the only one of the six Generation IV technologies that are cooled by water. Supercritical water-cooled reactors are designed to be a thermal reactor in the intermediate term, using enriched UO_2 as a fuel with a once-through fuel cycle. However, the ultimate goal is to build them as a fast neutron reactor with full actinide recycling. Very high-temperature gas reactors have an open fuel cycle. They will employ enriched UO_2 as a fuel, possibly in the form of pebbles coated with a graphite moderator like those required for PBMR.
- In the longer term, uranium resource availability could also become a limiting factor. Thus, one challenge to long-term, widespread deployment of Generation IV nuclear energy systems is that their fuel cycles must minimize the production of long-lived radioactive wastes while conserving uranium resources.
- Very high-temperature gas reactors, helium- and lead-cooled fast reactors and the molten salt reactor are all designed to generate electricity and also to operate at sufficiently high temperatures to produce hydrogen by thermochemical water cracking. Thermochemical hydrogen production can be achieved at temperatures of less than 900°C, using processes such as the sulfur–iodine cycle in which sulfur dioxide and iodine are added to water, resulting in an exothermic reaction that creates sulfuric acid and hydrogen iodide. At 450°C, the HI decomposes to iodine (which is recycled) and hydrogen. Sulfuric acid decomposes at 850°C, forming sulfur dioxide (which is recycled), water, and oxygen. The only feeds to the process are water and high-temperature heat, typically 900°C, and the only products are hydrogen, oxygen, and low-grade heat. Because of its potential to produce cheap and green hydrogen, VHTR technology has been given high priority by the U.S. DOE in its Next Generation Nuclear Plant (NGNP), which aims to make both electricity and hydrogen at very high levels of efficiency and near zero emissions.

Nuclear Energy Challenges

The main challenges nuclear power development faces in developed countries are economic and political, while the main issues in developing countries are a lack of adequate infrastructure, particularly in the back-end of the fuel cycle, a lack of expertise in nuclear technology and its safety culture, as well as financial issues. Developing countries are more likely to profit from the enhanced and passive safety features of the new generation of reactors, which have a stronger focus on the effective use of intrinsic characteristics, simplified plant design, and easy construction, operation, and maintenance. Public concerns about nuclear safety, national security, nuclear waste, and nonproliferation are also hindering the development of nuclear power. If nuclear power is to contribute in significant ways to meeting future energy demands, these issues, real or perceived, must be addressed.

Nuclear Safety. The International Nuclear Safety Advisory Group (INSAG) has suggested requiring that future nuclear plants be safer by a factor of 10 than the targets set for existing reactors (i.e., targets of 10^{-5}/year for core damage and 10^{-6}/year for large radioactive releases for future plants).[20]

There are four primary goals for safety issues in nuclear energy development:

- The first is reactivity control, which is the process of stopping fission reactions. The lack of reactivity control caused the Chernobyl reactor accident in 1986.
- The second safety goal is to reliably remove decay heat, which is the heat generated by radioactive decay of the fission products that continue to be produced even after fission reactions stop. Decay heat, if not removed, can result in overheating of and damage to the fuel, such as occurred in the Three Mile Island accident in Pennsylvania in March 1979.
- The third goal is to provide multiple barriers to contain the radioactive material. The barriers include the fuel cladding, the reactor vessel, and the containment building.
- The fourth goal is plant safety sufficient to eliminate the need for detailed evacuation plans, emergency equipment, and periodic emergency evacuation drills.

Although statistics comparing the safety of nuclear energy with alternative means of generating electricity show nuclear to be the safest, vast efforts to enhance the safety of advanced Generation III+ reactor designs and the revolutionary Generation IV technologies continue. These designs incorporate what are known as passive safety systems. Evolutionary designs explore many avenues to increased safety, including using modern control technology, simplifying safety systems, making use of passive designs, and extending the required response times for safety systems actuation and operator action. One example of an advanced reactor with passive safety systems is the economic simplified boiling-water reactor (ESBWR), which was developed by General Electric from its advanced boiling-water reactor design and is at the preapplication stage for NRC design certification. The Westinghouse AP1000 light-water reactor features advanced passive safety systems with fewer components than conventional PWRs. For example, compared with a conventional 1000-MW PWR, AP1000 has 50% fewer valves, 35% fewer pumps, 80% less pipe, and 85% less cable. The reactor has a modular design that will reduce construction time to as little as three years from the time the concrete is first poured to the time that fuel is loaded into the core.

Table 3 summarizes some key technical challenges and potential technical responses for a potential nuclear power increase in capacity while at the same time improving economic attractiveness.

However, in spite of the evolutionary improvements in safety, support for nuclear power has not increased in Western Europe and North America. One explanation is that, given there have been significant accomplishments in the area of safety, other issues (such as spent fuel and nuclear waste management) have replaced improvement of existing safety features as the greatest challenges to the future development of nuclear power.

Spent fuel and nuclear waste management. Nuclear energy produces both operational and decommissioning wastes, which must be contained and managed. Spent fuel can be safely stored for long periods in water-filled pools or dry facilities, some of which have been in operation for 30 years. Although the volumes of nuclear waste are small compared to waste from other forms of electricity generation, spent fuel and radioactive wastes from nuclear power plants and spent fuel reprocessing—and ultimately from plant decommissioning—still need to be managed safely. For now, most high-level waste from commercial nuclear power is either stored on-site or transported to interim storage sites. In fact, nuclear power is the only energy-producing industry that takes full responsibility for all its waste and figures these costs into its product, which is a key factor in sustainability. Desirable features of innovative nuclear fuel cycles are economic competitiveness; reduction of nuclear waste and the hazards associated

Table 3 Nuclear power development—key challenges and potential technical responses

Challenges	Availability of a technological response
Safety area eliminate severe core damage	Reactor and nuclear power plant designs based on maximum use of passive safety features
Waste management minimize the volume and toxicity of nuclear waste	Multirecycling of most dangerous long-lived radioactive elements in reactors with fast neutron spectrum
Nonproliferation furtherance of nonproliferation aims, namely that nuclear materials cannot be easily acquired or readily converted for nonpeaceful purposes	Recycling of fissile materials, together with radioactive minor actinides; integral nuclear fuel cycle without the stockpiling of Pu
Resource base increase the resource base more than tenfold	Multirecycling of plutonium in fast reactors

with its long-term storage; furtherance of nonproliferation aims, namely that nuclear materials cannot be easily acquired or readily converted for nonpeaceful purposes; and improved efficiency in resource use. Table 4 gives examples of recent work in innovative nuclear fuel cycles. Although no large-scale programs on innovative nuclear fuel cycles are being implemented at present, some countries are investigating the necessary steps for change in the current situation.

Experience with both storage and transport over half a century clearly shows that there is no technical problem in managing civil nuclear waste sans environmental impact. Shortage of capacity for spent fuel storage is today's eminent issue in several countries where long-term waste disposal policy remains unsettled. The greatest concern over the storage of high-level nuclear waste is that over such a long period of time, the containers in which waste is stored could eventually leak. However, the scientific and technical communities generally feel confident that technical solutions for spent fuel and nuclear waste conditioning and disposal already exist. The question has become politicized and focused on final disposal because high-level nuclear waste must be stored for thousands of years, but there is general consensus that geologic disposal of spent fuel or high-level radioactive waste from reprocessing can be carried out safely in stable, deep geologic formations. However, site selection remains a major political issue in most countries developing such facilities, and no such commercial facility has yet been authorized. Although most nations have identified potential underground storage sites and conducted geological and geophysical tests as to their suitability, no underground storage site has progressed beyond the planning stage. In the United States, which is perhaps the most advanced in the planning stage, President Bush authorized the construction of a nuclear waste repository at Yucca Mountain in Nevada in 2002.

Regardless of whether particular wastes remain a problem for centuries, millennia, or forever, there is a clear need to address the question of their safe disposal. Therefore, technological breakthroughs over a range of reactors and a range of reactor characteristics are now needed to cope with emerging issues, such as nonproliferation, environmental mitigation, economics, and enhanced safety and security needs.

ALTERNATIVE ENERGY PROSPECTS

It has been less than 100 years since fossil fuels became the predominant sources of energy in the world with the discovery of oil, the development of natural gas fields, and the widespread distribution of electricity from coal-powered central power plants. From the dawn of human civilization until about 100 years ago, human and animal muscle and wood, with lesser amounts of solar, wind,

Table 4 Innovative technologies related to nuclear fuel cycle

Attribute	Process and system	Relevant countries	Features
Fuel composition and process	Pyro-process	Japan, Russia, United States	Nuclear waste volume is smaller and process facility is simpler than those of wet process (expected economic and environmental advantages)
	Vibro-packed fuel	Russia, Switzerland	Fuel particle is directly produced from acid solution from reprocessing (economic merit is expected to compare to powder technology)
	DUPIC system	Canada, Rep. of Korea	Plutonium is not separated from PWR spent fuel (proliferation resistance is expected)
	Thorium fuel (Th–U, Th–Pu)	India, United States	Th resource is abundant. Fuel with Th–^{233}U composition generates less MA than U–Pu fuel
	Inert-matrix fuel	France, Japan, Switzerland	Due to chemically stable oxide, spent fuel is regarded as waste form (environmental mitigation)
Partitioning and transmutation (P–T) system	Accelerator driven system	France, Japan, United States	High neutron energy produced in accelerator destroys MA, LLFP. Sub-critical core enhances safety
	P–T system with fast reactor (FR)	Japan, Russia	Existing FR technology is applied for destruction of MA, LLFP
Fast reactor and fuel cycle system	Pb (+Bi) FR	Russia	Enhanced resource utilization, proliferation resistance, safety, and waste features

Source: From The National Academies Press (see Ref. 21).

hydro, and geothermal power, were the predominant sources of energy used by mankind. Can the renewable resources that sustained early civilization be harvested more cost effectively to meet a significant portion of the much higher demands of today's society?

Many regions of the world are rich in renewable resources. Winds in the United States contain energy equivalent to 40 times the amount of energy the nation uses. The total sunlight falling on the country is equivalent to 500 times America's energy demand. And accessible geothermal energy adds up to 15,000 times the national demand.[27] There are, of course, limits to how much of this potential can be used because of competing land uses, competing costs from other energy sources, and limits to the transmission systems needed to bring energy to end users. Moreover, the market penetration potential of renewable energy technologies will face a situation confronting any new technology and institutional constraints that attempt to resist an entrenched technology. But renewable energy technologies have made great progress in cost reduction, increased efficiency, and increased reliability in the past 30 years, as well as increasing contributions to the world's energy supply. In 2000, the global electricity capacity generated by renewable energy sources accounted for about 30% (102 GW) of the total electric power capacity (see Table 5).[2] Most of this energy came from hydroelectric power. However, other nonhydro renewable energy resources such as biomass, alcohol fuel, wind, geothermal, and solar energies have increased due to technological innovations and cost reductions and are becoming viable, commercially competitive, and environmentally preferable alternatives to traditional fossil fuels. While the global resources of fossil fuels gradually decline and environmental concerns associated with fossil fuels increase, the benefits of renewable energy are undeniably attractive.

Renewable energy sources have historically had a difficult time breaking into the existing markets dominated by traditional, large-scale, fossil fuel-based systems. This is partly because renewable and other new energy technologies are only now being mass-produced and have previously had high capital costs relative to more conventional sources because fossil fuel-based power systems have benefited from a range of subsidies, cheap raw materials, mass volume production over a span of many years, and a mature infrastructure. Moreover, the alternative energy sector has lacked appeal to investors in the United States because of heavy regulation, low growth, long-time return, and lack of support for innovative new companies in established energy markets.

The push to develop renewable and other clean energy technologies is no longer being driven solely by environmental concerns. Rather, these technologies are now also becoming economically competitive (Fig. 4). With continued research and deployment, many renewable-energy technologies are expected to continue their steep reductions in cost for such programs as wind, solar thermal, and photovoltaic energy. Several could become competitive over the next decade or two, either directly or in distributed utilities. Wind turbines, in particular, could even become broadly competitive with gas-fired combined-cycle systems within the next ten years in places where there are winds of medium or high quality. One disadvantage of renewable energy systems has been the intermittent nature of some sources, such as wind and solar energy. But this problem is not insurmountable—one solution is to develop diversified systems that maximize the contribution of renewable energy sources to meet daily needs, but also to use clean natural gas and/or biomass-based power generation to provide base-load power for the peak times in energy use, such as evening air conditioning or heating demand. Because of the changing U.S. electricity marketplace, remote or distributed markets for renewable electricity, as discussed above, appear to be more promising today than centralized electricity markets. Renewable energy sources are pinning their hopes on breakthroughs in the development of small, stationary and portable fuel cells and on the fast growing

Table 5 Renewable grid-based electricity generation capacity installed as of 2000

Technology	All countries (MW)	Developing countries (MW)
Total world electric power capacity	3400,000	150,000
Large hydropower	680,000	260,000
Small hydropower	43,000	25,000
Biomass power	32,000	17,000
Wind power	18,000	1,700
Geothermal power	8,500	3,900
Solar thermal power	350	0
Solar photovoltaic power (grid)	250	0
Total renewable power capacity	102,000	48,000

Source: From Renewable Energy Publications (see Ref. 2).

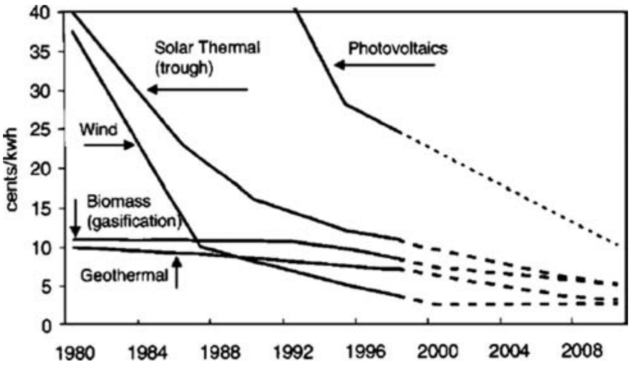


Fig. 4 The renewable electricity cost.
Source: From National Renewable Energy Laboratory (see Ref. 22).

market for them as minipower plants for use in factories, offices, retail stores, homes, and automobiles.

However, despite the significant benefits of alternative energy sources, according to the recent AEO2004 forecast, which assumes the world's oil supply peak will not occur before 2025, petroleum products are predicted to dominate energy use in the transportation sector. Energy demand for transportation is projected to grow from 26.8 quadrillion Btu in 2002 to 41.2 quadrillion Btu in 2025 (Fig. 5).

According to the forecast, motor gasoline use will increase by 1.8% per year from 2002 to 2025, when it will make up to 60% of transportation energy use. Alternative fuels are projected to displace only 136,800 barrels of oil equivalent per day in 2010 and 166,500 barrels per day in 2025 (2.1% of light-duty vehicle fuel consumption) in response to current environmental and energy legislation intended to reduce oil use. Gasoline's share of the demand is expected to be sustained by low gasoline prices and slower fuel efficiency gains for conventional light-duty vehicles (cars, vans, pickup trucks, and sport utility vehicles) than were achieved during the 1980s.

Therefore, even though renewable energy technologies have advanced dramatically during the past 30 years, fossil fuels seem slated to dominate the energy supply, and oil will remain the world's foremost source of primary energy consumption throughout the 2001–2025 period (at 39%) despite expectations that countries in many parts of the world will be switching from oil to natural gas and other fuels for their electricity generation. Robust growth in transportation energy use—overwhelmingly fueled by petroleum products—is expected to continue over the 24-year forecast period as shown in Fig. 6.[19] For this reason, oil is projected to retain its predominance in the global energy mix, notwithstanding increases in the penetration of new technologies, such as hydrogen-fueled vehicles.

The same trend of fossil fuel dominance over a slowly growing market in renewable energy can be found in the United States in the past five years. Renewable energy consumption in 2003 grew only 3% to 6.1 quadrillion Btu.[26] Overall, renewable energy contributed only 6% of the nation's total energy supply. Current levels of renewable energy use represent only a tiny fraction of what could be developed, but the U.S. energy infrastructure is mainly based on the consumption of fossil fuels. As shown in Fig. 7, solar and wind only accounted for 3% of total renewable energy use in 2003—less than 0.2% of the U.S. energy supply. Without government policies or programs—such as environmental laws aimed at limiting or reducing pollutants from the combustion of fossil fuel consumption and encouraging the use of nonfossil fuels—consumption of fossil fuels like oil, natural gas, and coal are expected to supply most of the primary energy needed to meet the projected demand for end-use consumption.

In 2001, the U.S. DOE produced a 50-year perspective for future U.S. Highway Energy Use.[4] The DOE assumed the demand for world oil products would grow at 2% per year. After 2020, when conventional oil production is assumed to peak—once 50% of ultimate resources have been produced and begin a continual decline—the gap

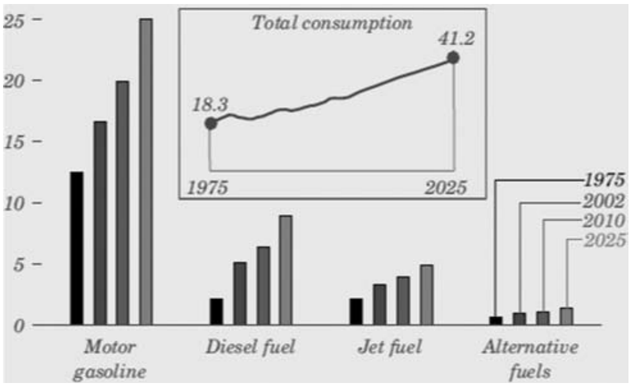


Fig. 5 Transportation energy consumption by fuel (quads). Source: From U.S. Department of Energy (see Ref. 6).

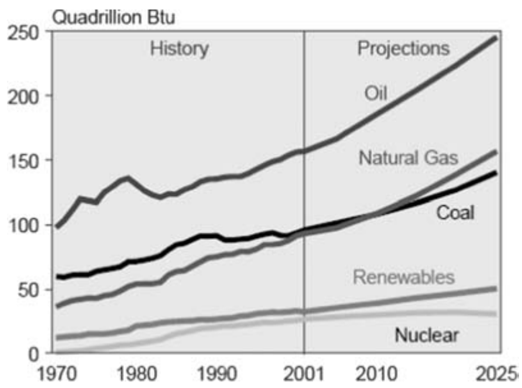


Fig. 6 World marketed energy consumption by energy source, 1970–2025.
Source: From Washington Times Corporation (see Ref. 25).

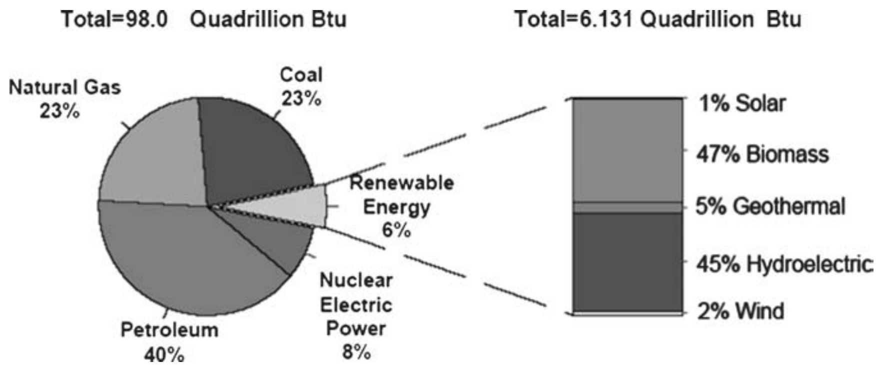


Fig. 7 Renewable energy consumption in the U.S.'s energy supply, 2003.
Source: From U.S. Department of Energy (see Ref. 26).

between continuing demand growth and declining production could be around the equivalent of 50 billion barrels of oil (145 MBPD) by 2050, or almost twice the current conventional oil production. Although there is considerable uncertainty, the report suggested that the United States should start transportation's energy transition immediately because the time needed to fully implement a new vehicle technology in all vehicles on the road will be 30 years or more, and fully implementing a new fuel will take even longer.

The market penetration of alternative renewable fuels, such as biofuel/biodiesel or hydrogen (see Table 6), are dependent on their cost-effective production, storage, delivery, safety, and customer acceptance, as well as related vehicle technology development, such as HEVs, battery-electric vehicles, and fuel cells.

Obviously, the transition to energy diversity with significant proportions of renewable energy and nuclear energy will change the world's use of energy, its economy, and the environment. A recent study on the costs and benefits of gradually increasing the share of the nation's electricity from wind, biomass, geothermal and solar energy, as proposed in renewable portfolio standards (RPS), range from a 4% reduction in carbon dioxide emissions in 2010 to a 20% reduction in 2020, which would freeze electricity-sector carbon dioxide emissions at year 2000 levels through 2020 at a modest cost of $18 per ton reduced.[27] In particular, the plains, western, and mid-Atlantic states are projected to generate more than 20% of their electricity from a diverse mix of renewable energy technologies. By contrast, carbon dioxide emissions are projected to grow 24% over the same period under a business-as-usual scenario. Renewable energy and energy efficiency are not only affordable, but their expanded use will also open new areas of innovation and business, creating opportunities and a fair marketplace for the transition to a clean energy economy, all of which will require leadership and vision from both government and industry. However, in the absence of government policies for both technology push and market pull, those renewable energy technologies will be not widely commercialized, and it will be difficult to extend the use of renewable sources on a large scale because most renewable energy sources are not expected to compete economically with fossil fuels in the mid-term if there are no significant changes in the cheaply-priced fossil fuel market.

CONCLUSIONS

Over the past 30 years, renewable energy technologies with unique environmental and social benefits have made significant progress as alternative energy options to fossil fuels. Of the many alternative sustainable energy technologies that have been developed, some are already making large inroads into the marketplace, such as nuclear power, biomass, hydropower, geothermal, wind, and solar power. Other technologies, perhaps those most beneficial to a sustainable future, require more efforts in research, development, and deployment before they can become economically viable and technically feasible. The recent rise in demand for energy, which is mostly due to increased demand in China and India as part of the expansion of their modern economy, coupled with the geopolitical aspects of the energy sector, has caused record high oil and gas prices across the globe. To confront the limited availability of what has been thus far viewed as inexpensive fossil fuels and growing energy demand from developing countries, the world is now necessarily on the brink of an energy transition from fossil fuels to clean renewable energy sources. However, the global energy infrastructure is mainly based on the consumption of fossil fuels, with very little being done to reduce dependence on these energy resources until now. Although after three decades of development, renewable energy has been proven technically and economically feasible, the timing of the transition will depend on government polices and the will of the customer. For example, current nuclear energy is technically and economically feasible to meet market demand, but certain social and safety issues remain that need to be addressed through a collaborative approach of

Table 6 Comparisons of gasoline and alternative transport fuels

	Gasoline	Biodiesel (B20)	Compressed natural gas (CNG)	Ethanol (E85)	Hydrogen
Chemical structure	C_4–C_{12}	Methyl esters of C_{16}–C_{18} fatty acids	CH_4	CH_3CH_2OH	H_2
Main fuel source	Crude oil	Soy bean oil, waste cooking oil, animal fats, and rapeseed oil	Underground reserves	Corn, grains, or agricultural waste	Natural gas, methanol, and other energy sources
Energy content per gallon	109,000–125,000 Btu	117,000–120,000 Btu (compared to diesel #2)	33,000–38,000 Btu @ 3,000 psi; 38,000–44,000 @ 3,600 psi	~80,000 Btu	Gas: ~6,500 Btu @ 3 kpsi; ~16,000 Btu @ 10 kpsi Liquid: ~30,500 Btu
Price per gallon/ gasoline gallon equivalent (GGE)	$1.99	$2.06	$1.40	$2.28	$5–$11 (2003) $1.5–$2.25 (2015)
Environmental impacts of burning fuel	Produces harmful emissions; however, gasoline and gasoline vehicles are rapidly improving and emissions are being reduced	Reduces particulate matter and global warming gas emissions compared to conventional diesel; however, NO_x emissions may be increased	CNG vehicles can demonstrate a reduction in ozone-forming emissions (CO and NO_x) compared to some conventional fuels; however, HC emissions may be increased	E-85 vehicles can demonstrate a 25% reduction in ozone-forming emissions (CO and NO_x) compared to reformulated gasoline	Zero regulated emissions for fuel cell-powered vehicles, and only NO_x emissions possible for internal combustion engines operating on hydrogen
Energy security impacts	Manufactured using imported oil, which is not an energy secure option	Biodiesel is domestically produced and has a fossil energy ratio of 3.3 to 1, which means that its fossil energy inputs are similar to those of petroleum	CNG is domestically produced. The United States has vast natural gas reserves	Ethanol is produced domestically and it is renewable	Hydrogen can help reduce U.S. dependence on foreign oil by being produced by renewable resources
Fuel availability	Available at all fueling stations	Available in bulk from an increasing number of suppliers. There are 22 states that have some biodiesel stations available to the public	More than 1,100 CNG stations can be found across the country. California has the highest concentration of CNG stations. Home fueling will be available in 2003	Most of the E-85 fueling stations are located in the Midwest, but in all, approximately 150 stations are available in 23 states	There are only a small number of hydrogen stations across the country. Most are available for private use only

(Continued)

Table 6 Comparisons of gasoline and alternative transport fuels *(Continued)*

	Gasoline	Biodiesel (B20)	Compressed natural gas (CNG)	Ethanol (E85)	Hydrogen
Maintenance issues		Hoses and seals may be affected with higher-percent blends, lubricity is improved over that of conventional diesel fuel	High-pressure tanks require periodic inspection and certification	Special lubricants may be required. Practices are very similar, if not identical, to those for conventionally fueled operations	When hydrogen is used in fuel cell applications, maintenance should be very minimal
Safety issues (Without exception, all alternative fuel vehicles must meet today's OEM safety standards)	Gasoline is a relatively safe fuel because people have learned to use it safely. Gasoline is not biodegradable though, so a spill could pollute soil and water	Less toxic and more biodegradable than conventional fuel, can be transported, delivered, and stored using the same equipment as for diesel fuel	Pressurized tanks have been designed to withstand severe impact, high external temperatures, and automotive environmental exposure	Ethanol can form an explosive vapor in fuel tanks. In accidents; however, ethanol is less dangerous than gasoline because its low evaporation speed keeps alcohol concentration in the air low and nonexplosive	Hydrogen has an excellent industrial safety record; codes and standards for consumer vehicle use are under development

Source: From U.S. Department of energy (see Ref. 24).

the world community. The recently rising prices of conventional sources of energy give the strongest economic argument in favor of the expanded market of renewable energy sooner rather than later. The evolution of any new energy technology will take a long time (more than 20 years) to be well accepted and established in the market. Energy substitution will begin in earnest when the costs of energy production by alternative methods are lower than the prevailing prices of conventional sources and when consumers are convinced there will be no reversal in price and supply trends. Although greater energy utilization efficiencies have proven to be effective conservation options in reducing energy consumption over past decades, they cannot and should not be viewed as an ultimate remedy or solution. In the long run, renewable energy and advanced nuclear energy seem to be the best options for meeting the increasing clean energy demand.

REFERENCES

1. Clemmer, S.L.; Donovan, D.; Nogee, A. *Clean Energy Blueprint: A Smarter National Energy Policy for Today and the Future*; Union of Concerned Scientists and Tellus Institute: Cambridge, MA, June 2001.
2. Martinot, E.; Chaurey, A.; Lew, D.; Moreira, J.R.; Wamukonya, N. Renewable energy markets in developing countries. Ann. Rev. Energy Env. **2002**, *27*, 309–348.
3. Zhang, M.; Eddy, C.; Deanda, K.; Finkelstein, M.; Picataggio, S. Metabolic engineering of a pentose metabolism pathway in ethanologenic zymomonas mobilis. Science **1995**, *267*, 240–243.
4. Birky, A. et al. Future U.S. highway energy use: a fifty year perspective (draft 5/3/01). *Office of Transportation Technologies, Energy Efficiency and Renewable Energy*; U.S. Department of Energy: Washington, DC.
5. Energy Information Administration. Annual Energy Review 1999, DOE/EIA-0384 (99), Washington, DC, 1999.
6. AEO2004 The Annual Energy Outlook 2004 (AEO2004). *Energy Information Administration, Office of Integrated Analysis and Forecasting*. U.S. Department of Energy: Washington, DC.
7. Energy Information Administration. *Alternatives to Traditional Transportation Fuels, 2003. Estimated Data*; Washington, DC, Febraury 2004.
8. Pearl, G.G. Animal fat potential for bioenergy use. *Bioenergy 2002*, The 10th Biennial Bioenergy Conference, Boise, ID, September 22–26, 2002.
9. EIA, Assumptions to the Annual Energy Outlook 2004, Report No: DOE/EIA-0554, February 2004.
10. Weiss, S.B. Can. J. Forest Res. **2000**, *30*, 1953.

11. DOE. Solar Energy Technologies Program—Multi-Year Technical Plan 2003–2007 and Beyond, Report No: DOE/GO-102004-1775, 2004.
12. Web site for the European Wind Energy Association, http://www.ewea.org/ (accessed May 19, 2004).
13. Schwartz, M.N.; Elliott, D.L. Arial Wind Resource Assessment of the United States. In *Alternative Fuels and the Environment*; Lewis Publishers: Boca Raton, FL, 1994; [Chapter 17].
14. DOE. Strategic Plan for DOE Hydrogen Program, DOE/GO-10 098-532, U.S. Dept. of Energy, 1998.
15. DOE. A National Version of America's Transition to A Hydrogen Economy–To 2003 and Beyond, National Hydrogen Version Meeting, Washington, DC, November 15–16, 2001.
16. IAEA, International Atomic Energy Agency, Reference Data Series 2, *Power Reactor Information System*, Web site www.iaea.org/programmes/a2/.
17. University of Chicago, *The Economic Future of Nuclear Power*, August 2004.
18. IEO. International energy outlook 2004. *Energy Information Administration, Office of Integrated Analysis and Forecasting*, U.S. Department of Energy: Washington, DC, April, 2004.
19. Nuclear Engineering International Handbook 2005.
20. Basic Safety Principles for Nuclear Power Plants (INSAG-12). 75-INSAG-3 Rev 1, IAEA, 1999.
21. Fukuda, K.; Bonne, A.; Mourogov, V.M. Global view on Nuclear Fuel Cycle—Challenges for the 21st century. Proceedings of the International Conference on Future Nuclear Systems Global'99, Jackson Hole, Wyoming, U.S.A.
22. Bull, S.R. Renewable energy today and tomorrow. Proc. IEEE **2001**, *89* (8), 1216–1226.
23. American Wind Energy Association, Global Wind Energy Market Report, Washington, DC, March 12, 2003.
24. DOE, The 12th Clean Cities Alternative Fuel Price Report, June 29, 2004 (http://www.eere.energy.gov/afdc/resources/pricereport/).
25. Tanenbaum, B.S. Fossil Fuels and Energy Independence, From The World &I, Washington Times Corporation, 2002; pp. 148–155.
26. Energy Information Administration (EIA), Renewable Energy Trends With Preliminary Data For 2003, 2004.
27. Clemmer, S.; Nogee, A.; Brower, M.C. A Powerful Opportunity: Making Renewable Electricity the Standard, Union of Concerned Scientists, November 1998.
28. DOE Hydrogen Program. The Hydrogen, Fuel Cells & Infrastructure Technologies Program Multi-Year Research, Development and Demonstration Plan, February 2005 (http://www.hydrogen.energy.gov/).
29. Generation IV International Forum (GIF) Online Reports, A Technology Roadmap for Generation IV Nuclear Energy Systems, December 2002 (http:gen-iv.ne.doe.gov/GENIVDOCUMENTs.asp).

ANSI/ASHRAE Standard 62.1-2004

Leonard A. Damiano
David S. Dougan
EBTRON, Inc., Loris, South Carolina, U.S.A.

Abstract
American National Standards Institute (ANSI)/American Society of Heating and Refrigeration and Air-Conditioning Engineers, Inc. (ASHRAE) Standard 62.1-2004 is a short but often misunderstood document outlining ventilation requirements intended to provide acceptable indoor air quality (IAQ) for new buildings or those with major renovations. Because of the rate-based nature of both procedures allowed for compliance, this analysis focuses on the practical needs of reliable intake rate control and the risks of indirect controls. Design recommendations offered are intended to increase the potential for both predictable compliance and the flexibility to accommodate future changes while providing the greatest control reliability with the most energy-efficient methods.

INTRODUCTION

This is a summary of American Society of Heating and Refrigeration and Air-Conditioning Engineers, Inc. (ASHRAE) Standard 62.1-2004, ventilation for acceptable indoor air quality (IAQ) in Commercial, Institutional, Industrial, and High rise residential buildings,[1] as it impacts and is influenced by ventilation control requirements, methods, and equipment. Operational implementation of these requirements can have a sizeable influence on energy usage when applied improperly or incompletely. Operational precision and design reliability are essential for energy minimization when compliance with 62.1 and energy codes are simultaneous goals.

This is not a condensed version and this does not cover all requirements of the Standard. Designers are strongly encouraged to read the entire 19-page document (a total of 44 pages with appendices). The Standard cannot be understood or properly applied without considering the relationships and interdependencies of requirements of the document.

This American National Standards Institute (ANSI)-approved standard has been developed by a Standing Standards Project Committee (SSPC) of the ASHRAE under a ‘continuous maintenance’ protocol. At any point in time, the ‘official’ Standard is comprised of both the most recently published parent document and all current addenda. The latest parent document was republished earlier this year to combine 17 addenda that had been approved subsequent to the original release of the 62-2001 ‘parent’ document in January 2002. The result is a final version that is substantially different from the basic ventilation standard we have used since 1989.[1]

In 2003, the scope of the Standard officially changed and a separate ASHRAE committee was formed to address the specific needs of low-rise residential buildings. The existing Standard became known as 62.1 and the new residential standard became 62.2.

The promised *62.1 User’s Manual* was recently published in December 2005. Work continues on a *Guideline 19P*, which is intended to provide design guidance for methods that exceed the minimum requirements of the Standard. Both of these supplemental documents should assist the designer and the facility operator in their understanding of and compliance with the Standard.

ANALYSIS AND RECOMMENDATIONS

Our discussion of Standard 62.1 will try to mimic the structure of the Standard, provide recommendations for compliance, and highlight methods and assumptions to avoid. Our objectives have determined the content.

The Standard’s “Purpose” and “Scope” are covered in Sections 1 and 2. To comply with the Standard, designers of mechanical ventilation systems are tasked to provide specific minimum rates of acceptable outdoor air to the breathing level of the occupied structures. In doing so, an acceptable indoor environment may be achieved providing improved occupant productivity and health. The procedures allowed for compliance with our national standard on ventilation are prescriptive or performance-based. Their selection and application should be evaluated for IAQ risk by the design practitioner.

Keywords: ASHRAE; Standard 62.1; Indoor air quality; IAQ; Ventilation; Airflow control; Building pressurization; Moisture management.

Definitions

Section 3 addresses the definition of terms used within the Standard. Noteworthy is the Standard's definition of "acceptable IAQ," which is defined as:

> …air in which there are no known contaminants at harmful concentrations as determined by cognizant authorities and with which a substantial majority (80% or more) of the people exposed do not express dissatisfaction.[1]

This means that 62.1, like all ASHRAE Standards, assumes that one out of five occupants (20%) might not be satisfied with the results of compliance and might express dissatisfaction with the IAQ, even if the Standard is followed perfectly. Many sources have concluded that the majority of Heating Ventilation Air Conditioning (HVAC) systems designed in the United States do not meet the minimum ventilation rates prescribed during operation. In which case, the actual occupant dissatisfaction level is exponentially greater in practice.[2] It is not uncommon for rates to fall below levels that result in occupant dissatisfaction significantly greater than 50%. Many systems cannot meet the minimum airflow requirements at the occupied space during operation because of design choices and equipment limitations or due to the dynamic nature of mechanical ventilation systems and the constant external forces acting on the building envelope.

The impacts from these continuously changing external conditions are not limited to variable air volume (VAV) systems.[2] Outdoor airflow rates will also vary for systems that provide a constant volume of supply air (CAV) to the conditioned space, as a result of:

1. Changes in wind or stack conditions on the intake system,[3]
2. Changes in filter loading, or
3. Changes in airflow requirements during an economizer cycle.

The lack of specific guidelines to overcome the effect of changing system dynamics on ventilation rates and air distribution for today's HVAC systems are partially to blame for many design deficiencies observed.

Unlike thermal comfort, the effect of IAQ is difficult to measure (The work item that caused the most controversy was an attempt to standardise design criteria for the indoor environment). The criteria developed in the process have been published as a CEN Technical Report CR 1752. It specifies the levels of temperature, air velocity, noise, and ventilation for occupied spaces. Values are given for three categories of environmental quality: A—a high level of expectation, B—a medium level and C—a moderate level. Supporting information is given on the derivations of the specified values of the parameters as well as to enable alternatives, such as different clothing levels, to be accommodated in the design assumptions. The most debatable section is on IAQ. Here, prominence is given to the evaluation of the required ventilation rate for comfort based on perceived air quality, the method developed by Professor Fanger and his colleagues in Denmark. While some data is presented, it is acknowledged that more research is needed to provide reliable information on pollution loads from materials and on the additive effects of emissions from multiple sources (Source: http://www.aivc.org/frameset/frameset.html?../Air/20_3/jackman.html~mainFrame accessed on June 2005). Many believe that the outdoor air levels specified by ASHRAE are too low and should actually be increased, as indicated by published research and reflected in European standards from CEN Technical Committee 156 and their publication CR1752.[4]

Outdoor Air Quality

Section 4 of the Standard describes a three-step process to evaluate outdoor air for acceptability. One of those steps requires examination of both the regional and the local air quality by the building owner. The section also specifies the documentation required to support the conclusions of this preliminary review.

If the outdoor air quality is found to be unsuitable per Section 4, then treatment may be required as indicated in Section 6.2.1. Outdoor air treatment involves the removal of the particulates and gases encountered that are in excess of the minimum standards cited by cognizant authorities in Section 4.1.

Systems and Equipment

Section 5 specifies the minimum systems and equipment required under Standard 62.1. Section 5.4 states:

> Mechanical ventilation systems shall include controls, manual or automatic, that enable the fan system to operate whenever the spaces served are occupied. The system shall be designed to *maintain the minimum outdoor airflow* as required by Section 6 *under any load condition*. Note: VAV systems with fixed outdoor air damper positions must comply with this requirement at minimum supply airflow.[1]

The Standard recognizes that changes in mixed air plenum pressure, up to 0.5 in. WG [125 Pa] variation on VAV systems, can significantly influence outdoor air intake flow rates. However, it neglects the significant influence of external pressure variations on all systems that result from changes in wind and stack pressures, which often exceeds 0.5 in. WG [125 Pa]. Therefore, providing the minimum outdoor airflow defined in Section 6 ‘effectively’ requires a dynamic control alternative for compliance—possibly the use of permanent devices capable of maintaining outdoor airflow rates.

Not mentioning airflow measurement is analogous to ignoring the requirement for temperature measuring devices to maintain continuous temperature control. Because many systems, especially VAV, have thermal load requirements that differ from their ventilation needs, the requirements of this section can be more sustainable if the multispace Eqs. 6-1 to 6-8 are calculated for the design supply flows to individual zones using the minimum outdoor air requirements to each zone. In order to achieve the industry “standard of care” in professional HVAC design, the mechanical engineer is required to determine which zones may become ‘critical’ and that the worst ‘critical zone’ is at its minimum supply airflow. Even with the average reduction potentials due to the new ventilation rate (Table 6-1), this could still impose a severe energy penalty to many VAV system designs.

Because of the requirements set forth in the Standard for compliance “under any load condition,” it is necessary to maintain a constant rate of outdoor airflow in dynamic systems. Logically, Section 5 should require continuous airflow measurement at the intake of all air-handling units with automatic controls that function to provide a building or space with a constant rate of outdoor air, regardless of the system size or type. Doing so would alleviate several practical issues, clarify application and compliance questions in Section 6.2.7, Dynamic Reset; Section 7.2.2, Air Balancing; and Section 8.4.1.8, Outdoor Airflow Verification. A continuous measurement requirement was explicitly stated in the draft Standard 62-89R before Standard 62 became politicized, which was much more complicated and vague to the point of confusion.

American Society of Heating and Refrigeration and Air-Conditioning Engineer’s new *62.1 User’s Manual* provides more insight into the appropriate, if not encouraged, use of permanent instrumentation for continuous measurement, by explaining further:

> For VAV systems that can operate under a wide range of operating conditions, the system must be designed to provide the minimum outdoor air rates under all reasonably anticipated operating conditions (see Ref. 5, pp. 6–33).
>
> To comply [with Std. 62.1], most VAV systems will need to be designed with outdoor airflow sensors and modulating dampers or injection fans (see Ref. 5, pp. 5–10). In most cases, an active control system must be provided at the air intake and sometimes at the zone level to ensure minimum rates are maintained… Note that a fixed-speed, outdoor air fan without control devices will not maintain rates within the required accuracy (see Ref. 5, pp. 5–11).

We are encouraged to use direct measurement feedback for continuous control on all VAV designs, even those using a powered outdoor air system (i.e., injection fan, heat recovery ventilators (HRV)/energy recovery ventilators (ERV), smaller dedicated outdoor air systems (DOAS), etc.). Although not contained in the society’s ‘minimum’ standard, ASHRAE is highlighting the potential source of problems and the more obvious means to avoid them.

We believe and recommend that Section 5 of the Standard should encourage the use of airflow measuring devices in the supply air to critical zones of VAV systems, allowing not only for improved operating savings for continuous verification of compliance and as a diagnostic tool but also to reset intake rates based on input from the continuous calculation of the multispace equations defined in Section 6.2. Although this may sound impractical to some designers, the technology is available and surprisingly cost effective, especially when considering the potential benefits in occupant productivity and health, reduced potential liability, and the energy savings available.

Pressurization and Mold

Addendum 62x (62-2001) was approved in 2004 for inclusion with the Standard, but problems identified immediately after publication generated addendum 62.1a (Table 1), which was just published in a supplement to the 2004 document during this writing (May 2006).

The proposed addendum only addresses positive pressure during periods of dehumidification.

Moisture is a prerequisite for mold and fungal growth and the condition should be avoided. Whenever the temperature of a building envelope is lower than the dew point of air migrating across it, there will be condensation and the potential for mold. Costs for design improvements or preventive actions prior to mold conditions can range from an additional \$1.50 to \$15/ft^2 [\$16–\$161/m^2]. Costs for mold remediation and repair can range from \$30 to \$65/ft^2 [\$323–\$700/m^2] [6]. That effectively equals a penalty for inaction ranging from 4 to 20 times that of the cost for prevention.

The *62.1 User’s Manual* also comments on pressure effects and ‘pressurization flow.’

> Positive pressure in hot, humid climates can also reduce interstitial moisture and resultant fungal microbial growth.

Table 1 Section 5.10

5.10 **Dehumidification Systems.** Mechanical air-conditioning systems with dehumidification capability shall be designed to comply with the following:

5.10.1 **Relative Humidity.** Occupied space relative humidity shall be designed to be limited to 65% or less at either of the two following design conditions:1) at the peak outdoor dew point design conditions and at the peak indoor design latent load, or when system performance is analyzed with outdoor air at the dehumidification design condition (that is, design dew point and mean coincident dry-bulb temperature) and with the space interior loads (both sensible and latent) at cooling design values and space solar loads at zero.

Note: System configuration and/or climatic conditions may adequately limit space relative humidity at these conditions without additional humidity-control devices. The specified conditions challenge the system dehumidification performance with high outdoor latent load and low space sensible heat ratio.

Exception: Spaces where process or occupancy requirements dictate higher humidity conditions, such as kitchens, hot tub rooms that contain heated standing water, refrigerated or frozen storage rooms and ice rinks, and/or spaces designed and constructed to manage moisture, such as shower rooms, pools and spas.

5.10.2 **Exfiltration.** For a building, the design minimum outdoor air intake shall be greater than the design maximum exhaust airflow when the mechanical airconditioning systems are dehumidifying.

Exception: Where excess exhaust is required by process considerations and approved by the authority having jurisdiction, such as in certain industrial facilities.

Note: Although individual zones within the building may be neutral or negative with respect to outdoors or to other zones, net positive mechanical intake airflow for the building as a whole reduces infiltration of untreated outdoor air. [1]

> Stack and wind effects can cause large regions, such as entire levels or façades, to be negatively pressurized.[6]
>
> A building that is excessively pressurized may cause damage to the structural integrity of the building envelope.[6]

Noting the proliferation of mold in buildings, the ASHRAE board issued *Minimizing Indoor Mold Through Management of Moisture in Building Systems* in June 2005, stating that sound moisture management should take precedence over energy cost savings.[7] This Position Paper outlines recommendations for the management of moisture in buildings by describing issues related to the topic and highlighting resources available through the society. This policy statement will filter through the society's organization and eventually impact technical programs, research, and standards.

Some of their recommendations for proper moisture management include:

- Building and system design, operation, and maintenance provide for drying of surfaces and materials prone to moisture accumulation under normal operating conditions.
- Mechanical system design should properly address ventilation air.
- The sequence of operation for the HVAC system should contain appropriate provisions to manage humidity, control pressurization, and monitor critical conditions.[7]

The flaw of ASHRAE's position is in ignoring the potential for high humidity alone to provide sufficient moisture content for mold growth. Studies have shown that in temperatures between 30 and 86°F [−1.1°C–30°C] with a minimum relative humidity of 70% RH (noncondensing infiltration), mold growth has appeared on plasterboard, brick, and concrete within 3 days. At 65.3°F [18.5°C] (with adequate RH and "inadequate" substrate), mold grows on building materials after 6 h. It was shown to take only 1 h to grow with "adequate" substrate.[2,8]

Part of the solution to preventing infiltration of unfiltered and unconditioned humid air appears to be simple. In 1996, the Florida Solar Energy Center first published a case study which identified that an extremely small negative pressure differential created conditions that lead to mold problems in a small commercial building. This was later supported by a 2002 ASHRAE journal article whose recommendations indicated that differential pressures as low as +0.004–+0.008 in. WG [1–2 Pa] will prevent moisture infiltration problems.[9] This counterflow overcomes most of the natural pressures that power moisture migration, namely vapor, temperature (stack), and wind pressures. Those periods when pressurization flow is insufficient to counter infiltration are generally limited in duration. Thereafter, the flow of air to the direction of higher dew point temperature can remove any residual moisture in the wall cavity.

Control precision and control stability should be key objectives when energy usage is to be minimized and dynamic control of space pressurization is used.

'Building pressure' is accomplished by creating a pressurization flow. Anything that changes the pressurization flow will result in fluctuations in building pressure. The pressurization flow is generally influenced by the HVAC system by controlling the volumetric differential of either the intake/relief air or the supply/return air. Heating ventilation air conditioning system control strategies that ignore these relationships or poorly implement them are widely known to have inherent pressurization problems.

Regardless of the system design, an effective method for maintaining pressurization flow is to monitor and control these airflow differentials. Typically, pressurization airflow (Q_P) is maintained at a fixed differential by an independent control loop; independent of the supply airflow rate required for temperature control, which uses a separate control loop. The airflow relationship is as follows:

$$Q_P = Q_{SA} - (Q_{RA} + Q_{EX(local)}), \text{ or}$$
$$Q_P = Q_{OA} - (Q_{EX(ahu)} + Q_{EX(local)})$$

where Q_P, pressurization airflow; Q_{SA}, supply airflow; Q_{RA}, return airflow; $Q_{EX(ahu)}$, exhaust/relief at AHU; $Q_{EX(local)}$, sum of local exhausts for zones served by AHU.[2]

The Standard must eventually address wind and stack effect and provide design guidelines that reflect conditions that influence buildings in their normal, native environment. In addition, increased humidity combined with wind- and stack-driven infiltration during periods when the ventilation system is not operating may be a significant factor influencing mold and fungal growth, e.g., offices and schools during closures. Designers and building operators should consider a limited night setback mode with provisions for humidity and pressurization flow controls. Such provisions would also tend to compensate for the building-generated contaminants by supplying a base ventilation rate, sufficient for minimal pressurization flow.

There is potential for condensation to occur under a positive pressure environment during periods of humidification in cold climates because the dew point of the air within a building could potentially be greater than the temperature of the building envelope. Maintaining a building at 'net neutral' pressure would be more appropriate under these conditions. 'Net neutral' control requires even more precise instrumentation and measurement directionality.

The widespread use of ERV in some geographic areas has decreased the amount of outdoor air available to pressurize a building and decreased the margin of error in control to maintain it. Although outdoor airflow rates into many buildings have increased with the use of ERVs, there is a strong potential for an increase in building pressurization problems which could lead to increased mold and fungal growth. Designers should exercise caution when implementing strategies that rely on ERV units for outdoor air and result in building design pressures that are close to net neutral. Wind, stack, and filter loading can easily result in depressurized buildings and increased condensation within the hidden cavities of the building envelope.[2]

Procedures

Section 6, Procedures, is the heart of the Standard. For compliance, designers must claim using either the ventilation rate procedure (VRP) or the indoor air quality procedure (IAQP) to determine the minimum dilution ventilation rate required for their design. Designers and operators cannot selectively ignore the parts they do not like. The entire procedure must apply. Parts from each procedure cannot be combined to achieve ventilation rates lower than those determined by the VRP alone. Great care should be given to the selection between these procedures.

The VRP, as defined in Section 6.1.1, "is a prescriptive procedure in which outdoor air intake rates are determined based on space type/application, occupancy level and floor area."[1] The key phrase is "prescriptive procedure in which outdoor air intake rates are determined." Very simply, this implies the need for some form of airflow measurement.

The alternative, IAQP in Section 6.1.2, "is a design procedure in which outdoor air intake rates and other system design parameters are based on an analysis of contaminant sources, contaminant concentration targets and *perceived* acceptability targets."[1] Any analysis of this procedure quickly reveals the clear discussion of airflow rate requirements based on varying contaminant levels.

Ventilation Rate Procedure

The VRP detailed in Section 6.2 is rate based. There is no question about this. In fact, the entire Standard is rate based, including the IAQP, which only provides the means to calculate allowable reductions in the design ventilation rate from those in Table 6-1 (see also Appendix D). The explicit statement to this effect was removed last year in an effort to render the language in the Standard more 'code enforceable.' Designers claiming compliance with the VRP must be able to document and substantiate that minimum intake rates are maintained during operation and "under all load conditions." They are not just 'capable of' maintaining—they are 'maintained' at no less than the higher of either code-required levels or those indicated by Table 6-1, and calculations in Section 6.2 of the Standard.

Once the outdoor air is determined to be acceptable or has been treated for use indoors, we can begin to determine how much is needed under our specific design situation. We can simplify the relationships and view the VRP with the aid of this partial flow chart from the new *62.1 User's Manual* (Fig. 1).

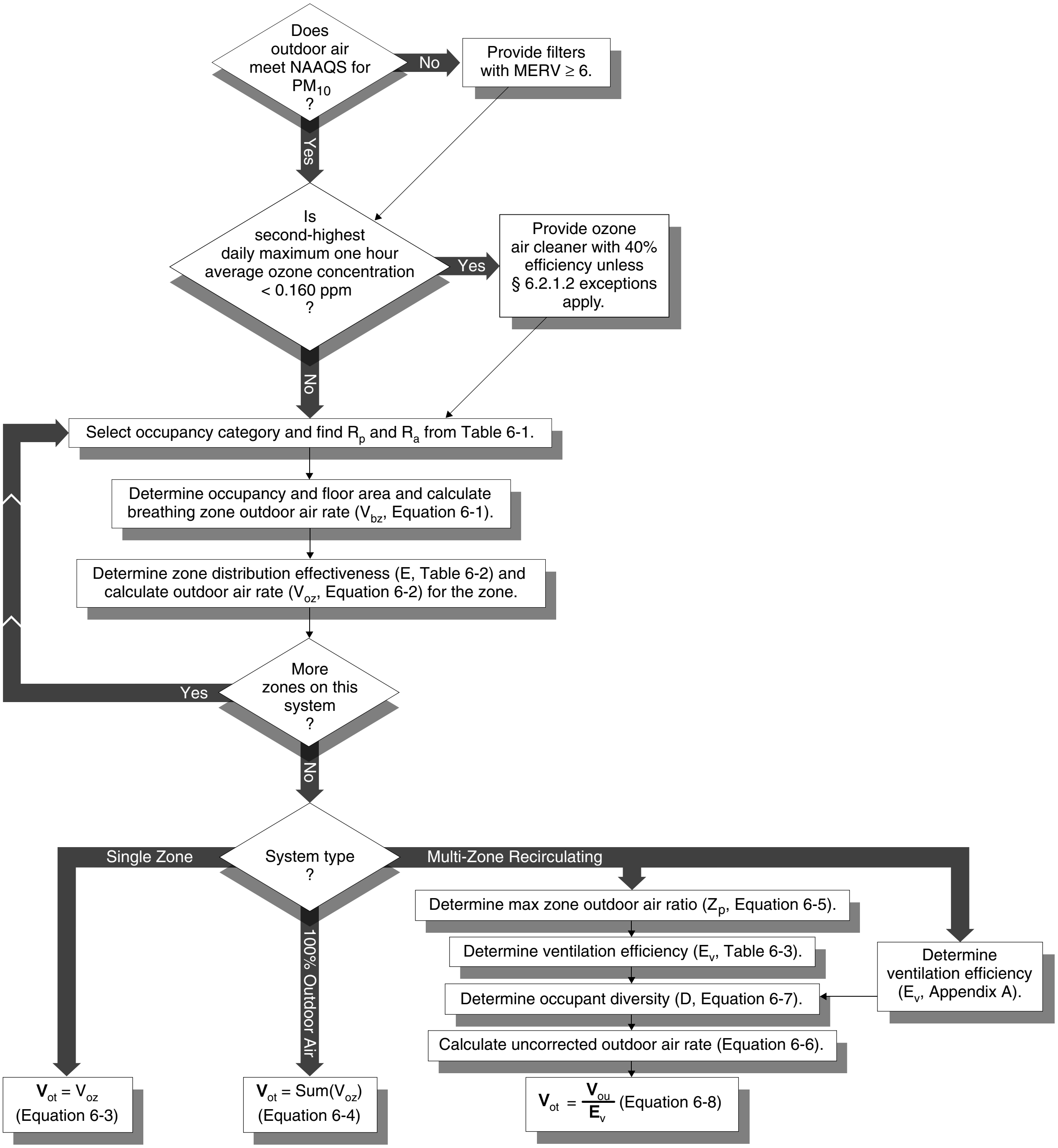


Fig. 1 Ventilation rate procedure (VRP) flow chart.
Source: From ASHRAE Standard 62.1 User's Manual (see Ref. 5).

First, we must calculate outdoor airflow requirements for the zone (V_{oz}), as detailed in Sections 6.2.2.1–6.2.2.3, which can be summarized with their corresponding equations and reference numbers below:

Calculate breathing − zone outdoor airflow:

$$V_{bz} = R_p P_z + R_a A_z \quad (1)$$

Determine zone air distribution effectiveness:

$$E_z = \text{Table 6-2}$$

Calculate zone outdoor airflow at diffusers:

$$V_{oz} = V_{bz}/E_z \quad (2)$$

Then, determine the outdoor airflow requirements for the system (V_{ot}) and calculate minimum outdoor air intake flow. We are given three general system types to choose from:

$$\text{Single}-\text{zone systems:} \quad V_{ot} = V_{oz} \tag{3}$$
$$100\%\ \text{OA systems:} \quad V_{ot} = \sum_{\text{all zones}} V_{oz} \tag{4}$$

and

$$\text{Multiple}-\text{zone recirculating systems}$$
$$\text{Outside Air Intake:} \quad V_{ot} = V_{ou}/E_v \tag{8}$$

In Multizone recirculation systems (V_{ou} and E_v), the variables needed to solve for V_{ot} are determined in Sections 6.2.5.1–6.2.5.4 and summarized below, but will be examined in more detail later.

Calculate the Zone primary outdoor air fraction:

$$Z_p = V_{oz}/V_{pz} \tag{5}$$

Determine the Uncorrected outdoor air intake:

$$V_{ou} = D\sum_{\text{all zones}} R_p P_z + \sum_{\text{all zones}} R_a A_z \tag{6}$$

and

Accounting for Occupant Diversity:

$$D = P_s/\sum_{\text{all zones}} P_z \tag{7}$$

One subtle change in the updated Standard includes the characterization, usage, and definition of "breathing zone." Plus, separate components are included to address both occupant and building-generated contaminants. Rates are no longer determined solely on occupancy "per person."

Total intake rates at the air handler can be directly determined with handheld instruments used in accordance with prescribed standards or by using an appropriate and permanently installed airflow measuring device. Total intake rates may be indirectly estimated by several other means (i.e., supply/return differential calculation, temperature balance, mass balance, steady-state CO_2 concentration, etc.). However, the uncertainty of indirect techniques introduces a significant level of risk (see Ref. 5, pp. 5–12).[2,10] The designer, facility owner, and occupants should carefully consider the method employed prior to implementation of any CO_2-based demand controlled ventilation (DCV) scheme as the sole method of intake rate determination.

The new VRP in Section 6.2 recognizes the magnitude of building-generated pollutants and subsequently added the "building component" in the zone ventilation equation. Table 6-1 and the accompanying notes specify outdoor air requirements for specific applications. Eq. 1 (Table 2) is now based on the combination of ventilation rates per person (as CFM/p) PLUS ventilation rates per floor area (as CFM/ft^2). Therefore, systems that meet these requirements:

1. The minimum requirements of Table 6-1 combined with
2. The calculated volume of outdoor air required by Section 6.2 and
3. The outdoor air quality requirements set forth in Section 4
4. While "under any load condition,"

can claim that their ventilation system complies with the Standard through the VRP.

Under ideal and very specific conditions, CO_2 levels can only reflect the rate that outdoor air enters the building on a per person basis, through any and all openings. Therefore, CO_2-based DCV with single 'ppm' set point control cannot be implemented under the new requirements of Standard 62.1 unless it is applied with excessive conservatism and the accompanying increase in energy usage. Otherwise, it will invariably underventilate spaces, overventilate spaces, or require that the Standard be rewritten or interpreted in such a way to allow the potentially large airflow errors that will result from using CO_2 sensor input (Table 3).

Calculated using the concentration balance formula in ASHRAE 62.1 Appendix C, at Various Population Densities in an Office Space[2]

Table 2 Steady-state CO_2 differentials

# People	Required total OA (CFM)	CFM/person	CO_2 rise [C_i–C_o] (ppm)	Comments
7 (+17%)	95	13.5	807	Underventilated
6 (base)	90	15	700	Using 700 ppm set point
5 (−17%)	85	17	644	Overventilated
3 (−50%)	75	26	438	

Area, 1000 ft^2; Total OA CFM required, 0.06 CFM/ft^2 + 5 CFM/person.
Source: From American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc. (see Ref. 1).

Table 3 Section 6.2.2.1

6.2.2.1 **Breathing Zone Outdoor Airflow.** The design outdoor airflow required in the *breathing zone* of the occupiable space or spaces in a *zone*, *i.e., the breathing zone outdoor airflow* (V_{bz}), shall be determined in accordance with Equation 6-1. $V_{bz} = R_p P_z + R_a A_z$ (6-1) where: A_z = *zone floor area*: the net occupiable floor area of the *zone* m^2, (ft^2). P_z = *zone population*: the largest number of people expected to occupy the zone during typical usage. If the number of people expected to occupy the zone fluctuates, P_z may be estimated based on averaging approaches described in Section 6.2.6.2. **Note**: If P_z cannot be accurately predicted during design, it shall be an estimated value based on the zone floor area and the default occupant density listed in Table 6-1. R_p = outdoor airflow rate required per person as determined from Table 6-1. **Note**: These values are based on adapted occupants. R_a = outdoor airflow rate required per unit area as determined from Table 6-1. **Note**: Equation 6-1 is the means of accounting for people-related sources and area-related sources for determining the outdoor air required at the *breathing zone.* The use of Equation 6-1 in the context of this standard does not necessarily imply that simple addition of sources can be applied to any other aspect of indoor air quality.[1]

The new *62.1 User's Manual* is really very well done. It will improve the readers' understanding of the Standard and provide needed design guidance with its examples. However, it does make a number of assertions that will leave you scratching your head. Here is one excerpt from the appendix on CO_2:

> This appendix describes how CO_2 concentration may be used to control *the occupant component* of the ventilation rate (see Ref. 6, pp. A–1).

This is a real problem for practitioners. The document intentionally leaves the details of how to address the 'building component' of the total rate required to the imagination of the designer wanting to use CO_2. It should advise the reader of at least the most correct method to use. If it cannot, the appendix should be deleted or it should not suggest that it offers guidance to "users."

It also appears that the fundamental basis for the Standard, maintaining minimum intake rates, has been transformed by the Appendix to the *User's Manual* to become the maintenance of "an acceptable bioeffluent concentration."[6] ASHRAE has published interpretations to the Standard, insisting that sensing for a single contaminant ignores potentially high levels of all other contaminants. To do so is to ignore all other design, environmental, or operational factors that impact the actual variable to be controlled—ventilation rates. Sensing an "indicator" of proper ventilation and using it for direct control raises liability issues and ignores what is really occurring at the outdoor air intake.

Indirect measurements for control typically carry such a large degree of uncertainty that one can never be secure that the controlled variable (ventilation rates) will not drop below or substantially exceed the mandated minimums under operating conditions.

There is too much risk. Compliance with the intelligent motor control (IMC) using CO_2 is questionable. Compliance with ASHRAE 62.1 is questionable. Why use it when there are other more reliable methods available to accomplish the same function or similar results?

Multispace 'Equations' Become a Design 'Procedure'

Once the breathing zone outdoor air requirement is determined, the Standard requires an adjustment based on the distribution system's efficiency and effectiveness. This makes complete sense because the air must reach the breathing zone to be effective.

Multizone recirculation systems are not as efficient as 100% OA systems and are therefore required to be factored by their approximate and relative inefficiency. We are given two methods to determine this factor:

1. Table 6-3, "default E_v" method
2. Appendix A, "calculated E_v" method

These methods produce significantly different results. The more precise method is contained in Appendix A, and as might be expected, it is more involved. The Table's conciseness requires it to be more conservative and therefore not as efficient in many situations.

For each multiple zone recirculating system (VAV or CAV), the primary outdoor airflow fraction must be calculated for all zones that may become 'critical' (only one zone can be critical on CAV systems). The "critical zone" is defined as the zone that has the highest percentage of outdoor air required in the primary air stream. When analyzing a VAV system dynamically, treat it as a CAV system.

As an example, if the supply air distribution system is located close to the return air, a short circuit is generally created. The Standard requires designers to use a zone air distribution effectiveness (E_v) of 0.5, which essentially doubles the amount of outdoor air required. In contrast to this example, a system with a ceiling supply and a ceiling return has a zone distribution effectiveness of 1.0 during cooling and 0.8 during heating. Therefore, the outdoor air setpoint must be reset seasonally or the more conservative factor used.

Systems that provide a variable supply of air volume to the conditioned space are influenced by everything previously discussed. In addition, outdoor airflow rates will vary as a result of changes in mixed air plenum pressure. If the design did not assume the worst-case scenario when the outdoor airflow rate for the air handler was determined, outdoor airflow rates on VAV systems may need to be reset based on calculations of the multispace equations (Eqs. 6-5 to 6-8, defined in Table 4), in order to avoid potentially excessive over ventilation and the associated energy penalty.

Advanced VAV control strategies can satisfy the requirements of Sections 6.2.5.1–6.2.5.4 dynamically and therefore more efficiently than static strategies. This can be accomplished by automatically determining the critical zone fraction to continuously calculate the corrected fraction of outdoor air. The calculation requires that the total supply airflow rate be continuously measured and that the airflow rate of the critical zones is measured with permanent airflow measuring devices capable of accurate measurement.

Airflow sensors provided with VAV boxes should not be used for these calculations. Although the original equipment manufacturer (OEM) devices may be adequate in modulating a terminal box for thermal comfort, the combination of typically poor inlet conditions, low quality airflow pickups, and low cost pressure sensors in the direct digital control (DDC) controller will not result in the measurement accuracy necessary for proper calculation of Eqs. 6-1 and 6-5 to 6-8. Conservative mathematical modeling has demonstrated that typical VAV box measurement performance

Table 4 Section 6.2.5

6.2.5 **Multiple-Zone Recirculating Systems.** When one air handler supplies a mixture of outdoor air and recirculated return air to more than one zone, the *outdoor air intake flow* (V_{ot}) shall be determined in accordance with section 6.2.5.1 through 6.2.5.4 [Equations 6-5 through 6-8].

6.2.5.1 **Primary Outdoor Air Fraction.** When Table 6-3 is used to determine system ventilation efficiency, the *zone primary outdoor air fraction* (Z_p) shall be determined in accordance with Equation 6-5.

$$Z_p = V_{oz} / V_{pz} \qquad (6\text{-}5)$$

where V_{pz} is the zone primary airflow, i.e., the primary airflow to the zone from the air handler including outdoor air and recirculated return air.
Note: For VAV systems, V_{pz} is the minimum expected primary airflow for design purposes.

6.2.5.2 **System Ventilation Efficiency.** The *system ventilation efficiency* (E_v) shall be determined using Table 6-3 or Appendix A.

6.2.5.3 **Uncorrected Outdoor Air Intake.** The design *uncorrected outdoor air intake* (V_{ou}) shall be determined in accordance with Equation 6-6.

$$V_{ou} = D \ \textit{all zones}\ R_p P_z + \ \textit{all zones}\ R_a A_z \qquad (6\text{-}6)$$

The *occupant diversity*, *D*, may be used to account for variations in occupancy within the zones served by the system. The *occupancy diversity* is defined as:

$$D = P_s / \textit{all zones}\ P_z \qquad (6\text{-}7)$$

where the *system population* (P_s) is the total population in the area served by the system.
Alternative methods may be used to account for population diversity when calculating V_{ou}, provided that the resulting value is no less than that determined by Equation 6-6.
Note: The *uncorrected outdoor air intake* (V_{ou}) is adjusted for diversity but uncorrected for ventilation efficiency.

6.2.5.4 **Outdoor Air Intake.** The design *outdoor air intake flow* (V_{ot}) shall be determined in accordance with Equation 6-8.

$$V_{ot} = V_{ou} / E_v \qquad (6\text{-}8)\text{”}$$ [1]

can be statistically exceeded by boxes without a measurement device.[2] Accurate airflow measuring devices having a total installed accuracy better than 5% of reading at maximum system turndown should be installed in the supply ducts for critical zones.

Increased precision allows terminal box selection for optimum energy performance as well as improved sound performance. Accordingly, a recently proposed ASHRAE TC1.4 Research Work Statement claims that using VAV box sensors and controllers allowing "a 20% minimum airflow setpoint can save $0.30/ft^2-yr. [in energy costs]. Multiplied across the millions [billions?] of square feet of commercial space served by VAV boxes, the potential economic, health and productivity benefits are significant."

Consider, for example, 1.5 million ft^2 of occupied area in a single 50+ story Manhattan high-rise office building. The savings potential equates to $450,000 per year in energy for one building. The 1999 Commercial Buildings Energy Consumption Survey (CBECS) showed 12 of 67.3 billion total ft^2 of commercial space as offices (2003 Report by Iowa Energy Center, 1999 CBECS, http://www.eia.doe.gov/emeu/efficiency/cbecstrends/pdf/cbecs_trends_5a.pdf). If we assume that 28% ((using 70% of 40% VAV offices) http://www.buildingcontrols.org/download/untracked/NBCIP_PNWD_3247%20F.pdf accessed July 2005 of VAV office buildings are capable of being upgraded, the United States would be looking at a conservative (50% of $0.30/ft^2) potential annual energy savings of $250 million per year. With total commercial space projected by department of energy (DOE) to increase 46% to 105 billion ft^2 by 2025 and all future VAV office designs capable of saving $0.30/ft^2, yes, the potential is understatedly "significant."

The results from these multispace equations can provide wide variations in outdoor airflow requirements in some systems. Increasing the critical zone supply flow while using reheat can reduce total outdoor airflow rates and overall energy usage. This method has been simulated using the multispace equation from Standard 62-2001 at Penn State University, with published results showing greater energy efficiency than the same system supplying the maximum, worst-case V_{ot} continuously. The basic variables, relationships, and end results should be the same using the VRP of Standard 62.1.

Then, the VRP continues and provides us with additional options to help make the design more specific to the designer's needs and to the demands of the situation. You may…

- Design using the short-term "averaged" population rather than the peak—Table 5 (9)
- Operate (and dynamically reset requirements) using "current" population data—Table 6, "DCV"

The peak population value may be used as the design value for P_z. Alternatively, time-averaged population determined as described in Section 6.2.6.2 may be used to determine P_z.

Outdoor airflow rates can also be reduced if the critical zones have variable occupancy or other unpredictably variable (dynamic) conditions. Changes in occupancy (or ventilation 'demand') can be detected in many ways, as indicated in the 'note' below. Therefore, DCV systems functioning to dynamically adjust the outdoor air intake setpoint should not be limited to CO_2 measurement input alone.

Section 6.2.7 (Table 6) on Dynamic Reset addresses conditions when the ventilation control system…

Table 5 Section 6.2.6

6.2.6 **Design for Varying Operating Conditions.**
6.2.6.1 **Variable Load Conditions.** Ventilation systems shall be designed to be capable of providing the required ventilation rates in the breathing zone whenever the zones served by the system are occupied, including all full- and part-load conditions.[1]

Table 6 Section 6.2.7

"…may be designed to reset the design *outdoor air intake flow* (V_{ot}) and/or space or zone airflow as operating conditions change. These conditions include but are not limited to:

1. Variations in occupancy or ventilation airflow in one or more individual zones for which ventilation airflow requirements will be reset.
   Note: Examples of measures for estimating such variations include: occupancy scheduled by time-of-day, a direct count of occupants, or an estimate of occupancy or ventilation rate per person using occupancy sensors such as those based on indoor CO_2 concentrations.
2. Variations in the efficiency with which outdoor air is distributed to the occupants under different ventilation system airflows and temperatures.
3. A higher fraction of outdoor air in the air supply due to intake of additional outdoor air for free cooling or exhaust air makeup.[1]

There is no indication in the Standard of how to implement Dynamic Reset with CO_2, which was left to be addressed by the *User's Manual*. The *User's Manual* included an appendix intended to address the continuing questions by users and designers: "How can we apply CO_2 measurements to DCV and comply with the VRP of Standard 62.1?" To this end, the appendix hedged sufficiently to avoid answering this question directly, underscoring the potential problems in applying DCV and simultaneously employing outdoor airflow measurement.

This section (Section 6.2.7), independent of the *User's Manual*, would lead one to believe that CO_2 is a method of "counting" and not an input to be used for direct ventilation control. Any counting method can be used to reset a flow rate established and controlled by some other measurement means. Intake rates should not be indirectly determined by the "counting" method.

Indoor Air Quality Procedure

Section 6.3 Indoor Air Quality Procedure begins…

> The Indoor Air Quality Procedure is a performance-based design approach in which the building and its ventilation system are designed to maintain the concentrations of specific contaminants at or below certain limits identified during the building design and to achieve the design target level of *perceived* IAQ *acceptability* by building occupants and/or visitors….[1]

The concept of providing "performance-based" solutions is desirable in principle. However, there are numerous risks associated with both the quantitative and subjective evaluations provided within the IAQ procedure that every designer should understand.

Because there are numerous contaminants that either will not be detected or for which "definite limits have not been set," this portion of the procedure has significant risks associated with it. It is unlikely that all contaminants of concern will be evaluated or reduced to acceptable levels. It is also not practical to measure all potential contaminants, and in some cases, such as with fungus or mold, measurement may not be possible.

Table 7 combined with Table 8 (b) and (c) emphasize the risk associated with the Indoor Air Quality procedure. The uncertainty of using "subjective occupant evaluations" together with the admittedly limited listings in Appendix B-SUMMARY OF SELECTED AIR QUALITY GUIDELINES, may be too great for many designers to 'claim' this procedure for compliance.

However, the concept of controlling the source of the contaminants makes perfect sense and is, in our opinion, more properly utilized under design approach Table 8 (d).

Because airflow rates are typically reduced in the IAQP, the measurement and control of intake rates are even more critical, especially on systems where the thermal load change is independent of the occupants and their activities. In addition, caution should be exercised when reducing outdoor airflow rates because it is also required to maintain proper building pressure, helps to minimize energy use, improves comfort control, and prevents mold growth within wall cavities.

Design Documentation Procedures

Section 6.4, Design Documentation Procedures, states:

> Design criteria and assumptions shall be documented and should be made available for operation of the system

Table 7 Section 6.3.1.3

6.3.1.3 **Perceived Indoor Air Quality.** The criteria to achieve the design level of acceptability shall be specified in terms of the percentage of building occupants and/or visitors expressing satisfaction with perceived indoor air quality."

Table 8 Section 6.3.1.4

6.3.1.4 **Design Approaches.** Select one or a combination of the following design approaches to determine minimum space and system outdoor airflow rates and all other design parameters deemed relevant (e.g., air cleaning efficiencies and supply airflow rates).

(a) Mass balance analysis. The steady-state equations in Appendix D, which describe the impact of air cleaning on outdoor air and recirculation rates, may be used as part of a mass balance analysis for ventilation systems serving a single space.

(b) Design approaches that have proved successful in similar buildings...

(c) Approaches validated by contaminant monitoring and subjective occupant evaluations in the completed building. An acceptable approach to subjective evaluation is presented in Appendix B, which may be used to validate the acceptability of perceived air quality in the completed building.[1]

(d) Application of one of the preceding design approaches (a,b, or c) to specific contaminants and the Ventilation Rate Procedure would be used to determine the design ventilation rate of the space and the IAQ Procedure would be used to address the control of the specific contaminants through air cleaning or some other means.[1]

> within a reasonable time after installation. See Sections 4.3, 5.2.3, 5.17.4 and 6.3.2 regarding *assumptions that should be detailed* in the documentation.[1]

Within Section 5.2.3, Ventilation Air Distribution requires us to:

> …specify minimum requirements for air balance testing or reference applicable national standards for measurement and balancing airflow.[1]

Providing permanently installed instruments and controls that result in and verify compliance with ASHRAE Standard 62.1 is perhaps one of the best reasons to provide such devices as part of any HVAC system design. Continuous data inputs may also be used to aid start-up, test and balance, commissioning, and measurement and verification (M&V) for energy usage and ongoing diagnostics. More precise and more reliable control could be viewed as a bonus.

Construction and System Start-Up

Section 7 addresses the construction and start-up phases of the project and has been included because a significant number of documented IAQ cases were a result of activities which took place during these phases of the project. The construction phase, addressed in Section 7.1 of the Standard, applies to "ventilation systems and the spaces they serve in new buildings and additions to or alterations in existing buildings."[1] The Standard addresses both the protection of materials and protection of occupied areas.

Mechanical barriers are specified to protect occupied areas from construction-generated contaminants. In addition, the HVAC system must be able to maintain occupied spaces at positive pressures with respect to the construction areas. In many cases, the HVAC system does not have the adequate capacity and controls to provide a barrier to the migration of contaminants using positive pressurization flow. Designers must consider the condition of the existing ventilation system and its ability to maintain a pressurized environment for spaces expected to continue occupancy, prior to initiating physical construction activities at the site.

The start-up phase, covered in Section 7.2, provides guidelines for air balancing, testing of drain pans, ventilation system start-up, testing of damper controls, and documentation requirements.

Section 7.2.2, Air Balancing, requires that systems be balanced, "at least to the extent necessary to verify conformance with the total outdoor air flow and space supply air flow requirements of this standard."[1] Unfortunately, the airflow rates of the system will vary after this activity has occurred, in most systems, for reasons discussed in the analysis of Section 5, Systems and Equipment. When applied in accordance with the manufacturer's recommendations, some airflow measuring devices only require the verification of operation by test and balance professionals. This TAB "snapshot" of airflow rates is analogous to providing a one-time setup for temperature control, which obviously would not be very effective. Providing permanently mounted airflow measuring stations would also support compliance with and reduce the time required to supply the documentary requirements for ventilation set forth in Section 7.2.6 (c).

Operations and Maintenance

All systems constructed or renovated after the date the Standard's 2001 version parent document was originally adopted are required to be operated and maintained in accordance with the provisions set forth in these newer sections of the Standard. It is important to recognize that if the building is altered or if its use is changed, the ventilation system must be reevaluated. Buildings that are likely to be changed or altered during their life spans should consider including a robust HVAC system design that takes into account changes in airflow rate requirements imposed by this Standard. Of course, provisions for permanently mounted airflow measurement devices and controls would significantly reduce both the cost and time associated with such changes as long as the HVAC load capacity could accommodate future requirements.

Section 8.4.1.7 addresses sensors. "Sensors whose primary function is dynamic minimum outdoor air control, such as flow stations…"[1] is discussed in this section even though they were not mentioned under Section 5, Systems and Equipment. Section 8.4.1.7 requires that sensors have their accuracy verified "once every six months or periodically in accordance with the Operations and Maintenance Manual."[1] The Operations and Maintenance Manual for some airflow measuring devices does not recommend periodic recalibration. Permanently calibrated airflow instrumentation has a significant advantage over other airflow measuring technologies and CO_2 sensors, whose transmitters are subject to frequent adjustments, zeroing, or regular calibrations to correct for analog electronic circuitry and sensor drift.

However, Section 8.4.1.8 (Outdoor Air Flow Verification) only requires the verification of airflow rates "once every five years."[1] Because external and system factors change continuously, clearly influencing outdoor airflow rates, this requirement does little to assure that proper ventilation rates are maintained under normal operation at different times of the day and the year. It effectively places the burden of verification of new building/system performance on the building operator, who is often not in a position to make such a determination.

This apparent contradiction with Section 8.4.1.7 will likely be examined by the ASHRAE SSPC62.1 committee in the near future. Permanent outdoor airflow measuring stations would provide continuous verification and the necessary control inputs to maintain ventilation requirements, automatically minimizing intake rates for energy usage and preventing other control inputs from causing a maximum intake limit from being exceeded.

CONCLUSIONS

American Society of Heating and Refrigeration and Air-Conditioning Engineers Standard 62.1 prescribes ventilation rates for acceptable IAQ. It should be clear to the building operator and the design professional that the dynamic nature of mechanical ventilation requires dynamic control to insure the continuous maintenance of specific predetermined conditions. As a rate-based standard, continuous airflow measurement should logically be a central component of any effective control strategy to assure acceptable IAQ and minimize the costs of energy to provide it.

REFERENCES

1. ASHRAE Standard 62.1-2004. *Ventilation for Acceptable Indoor Air Quality in Commercial, Institutional, Industrial and High Rise Residential Buildings*; American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc.: Atlanta, GA, 2005.
2. Dougan, D.S. IAQ by design and airflow measurement for acceptable indoor air quality. *Proceedings from IAQ Seminars*; EBTRON, Inc.: Loris, SC, 1999–2005.
3. Solberg, D.P.W.; Dougan, D.S.; Damiano, L.A. Measurement for the control of fresh air intake. ASHRAE J. **1990**, *32* (1), 46–51.
4. CEN (European Committee for Standardisation) Design criteria and the indoor environment. In *The Technical Report CR1752*; TC156 Ventilation for Buildings: Brussels, 1998–2004.
5. ASHRAE. 62.1 *User's Manual*; American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc.: Atlanta, GA, 2005; Dec 2005.
6. Bailey, R. Personal Correspondence. Chair ASHRAE TC1.12 Moisture Management in Buildings, Bailey Engineering: Jupiter, FL, April 13, 2005.
7. ASHRAE Board of Directors. Minimizing Indoor Mold Through Management of Moisture in Building Systems. Position Paper, American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc.: Atlanta, GA, June 2005.
8. Krus, M.; Sedlbauer, K.; Zillig, W.; K̈nzel, H.M. *A New Model for Mould Prediction and Its Application on a Test Roof*; Fraunhofer Institute for Building Physics, Branch Institute Holzkirchen: Munchen, 1999; (www.hoki.ibp.fhg.de/ibp/publikationen/konferenzbeitraege/pub1_39.pdf accessed March 2005).
9. Brennan, T.; Cummings, J.; Lstiburek, J. Unplanned airflows and moisture problems. ASHRAE J. **2002**, *44* (11), 48.
10. Krarti, M.; Brandemuehi, M.J.; Schroeder, C.; Jeannet, E. *Techniques for measuring and controlling outside air intake rates in variable air volume systems*, Final Report on ASHRAE RP-980; University of Colorado: Boulder, CO, 1999.

Auditing: Facility Energy Use

Warren M. Heffington
Industrial Assessment Center, Department of Mechanical Engineering, Texas A&M University, College Station, Texas, U.S.A.

Abstract
This entry discusses the energy audit and assessment processes, including the analysis of utility data, walk-through assessments, detailed assessments, and the reporting of results. This article will also provide a brief history of energy auditing, as well as a look toward the future.

INTRODUCTION

Energy assessments are associated with energy conservation programs and energy efficiency improvements. However, cost containment (usually associated with utility cost savings) is always the most sought after result. Changes that reduce operating expenses sometimes save little energy (e.g., correcting a power factor) or may increase energy use (e.g., replacing purchased fuel with opportunity fuel). Often known as energy audits, an older term that is being replaced by "energy assessments," these analyses of facility or system performance are aimed at improving performance and containing operating costs. Facilities that may be audited for energy use can be commercial or institutional buildings, industrial plants, and residences; examples of the systems that may be audited include air conditioning, pumping, and industrial production lines. The end of this article contains important references on the subject of energy assessment.[1,2]

Energy assessments today owe much to the National Energy Conservation Policy Act of 1978. This act resulted in the Institutional Conservation Program, a nationwide program of the U.S. Department of Energy (DOE) which called for matching grant funding to implement energy efficiency retrofits in secondary schools and hospitals. Directed by state energy offices, energy assessments predicted savings from these retrofits. A corresponding national DOE program is the Industrial Assessment Center (IAC) program,[3] which provides assessments of manufacturing plants and also educates university students about energy conservation. Sometimes pollution prevention, waste reduction, and productivity improvement projects are included with the energy assessments because the techniques for analyzing such projects are similar. Other DOE-sponsored assessment programs include the Federal Energy Management Program[4] for federal facilities and plant-wide assessments for private industries have also led to informative case studies and energy-efficient technology transfer.

The following section discusses the process of conducting an energy assessment, including pre-assessment activities, the actual assessment, and reporting of findings.

ASSESSMENTS

The energy assessment process may be conducted by a single user of a relatively unsophisticated software package that focuses on a few simple energy issues or a team of professional engineers who analyze problems using sophisticated software and complex calculations. The individual energy assessor might conduct assessments of residences as a city or utility service, and provide simple printouts of recommended changes for homeowners with some savings data and perhaps cost estimates. The more expensive and detailed assessments by teams of engineers may be reserved for large buildings and industrial complexes. In those cases, the feedback is often in the form of a formal technical report containing conceptual designs that allow capital decisions to be evaluated and made about proceeding with further analysis and planning.

The main objective of all assessments is to identify, quantify, and report on cost containment projects that can be implemented in a facility or system. These projects go by several names: energy conservation opportunity (ECO), energy conservation measure (ECM), energy cost reduction opportunity (ECRO), assessment recommendation (AR), and others. These projects are the heart of any energy analysis.

Assessments are characterized by the level of effort required and by the capital necessary to accomplish the recommended measures. The simplest type is a walk-through or scoping assessment and may result in no more than a list of possible projects for consideration. More complex, detailed assessments result in a formal technical report showing calculations of savings and

Keywords: Energy audits; Energy assessments; Energy conservation; Demand-side management; Energy audit reporting; Energy assessment reporting.

implementation costs. Walkthroughs and detailed assessments may be subdivided into projects requiring little capital and capital intensive projects. These low-capital projects are sometimes tasks that employees such as maintenance personnel should be accomplishing as part of their regular duties, sometimes known as maintenance and operation (M&O) opportunities,. ASHRAE designates the assessments and projects by levels: Level I—Walkthrough Assessment, Level II—Energy Survey and Analysis, and Level III—Detailed Analysis of Capital Intensive Modifications.[5] The most common measure of financial merit is simple payback although life cycle costing may be used and is required by some federal programs under the Federal Energy Management Program.[6]

ASSESSMENT PROCEDURE

Pre-Assessment Activities

Pre-assessment visit activities allow the assessment team to learn about the facility prior to an actual visit and should include obtaining the facility's energy consumption history. Often important clues about cost-containment projects come from historical information available from the facility or system operators' records for all significant energy streams. All assessments should be preceded by a review of at least 12 consecutive months of bills. In the case of energy, both energy in common units such as kilowatt hour (kWh), thousand cubic feet (MCF), or million British thermal unit (Btu) and cost in dollars should be graphed so that fluctuations in usages and costs are visible. The American National Standards Institute provides examples of these graphs.[7] Electrical demand in kilowatt (kW) or kilovolt ampere (kVA) should be plotted also, as should load factors and energy use, and energy cost indices in the case of buildings. Cost savings measures that can be identified from this review may result from inappropriate tariffs and tax charges, for example. This information gives an indication of possible projects to pursue during future phases of the assessment.

If pollution and waste are to be considered, such things as annual waste summaries and pollution information, as required by government regulators under the Clean Air Act or other legislation, should be reviewed. Water bills contain information about sewer charges as well as water consumption. Obtaining the tariff schedules for the various charges are important and all monthly charges should be recomputed to assure there are no mistakes. Tariffs in general are not well understood by users, who may not have the time or personnel to review the charges for errors. There are two important results of recalculation. First, the reviewer can reassure the client that the charges are correct (this can be comforting to most clients who have no good idea about the correctness of charges) and the reviewer can identify the rare error.

Second, facility energy bills and associated tariffs provide information so that the assessor can develop the avoided cost of energy (and demand) applicable to each billing account. The avoided cost usually is not evident directly in the tariff and typically the units for avoided costs are the same as the units of pricing. For example, the avoided cost of electrical energy measured in $/kWh may contain the block cost, fuel adjustment, and perhaps other charges. Sometimes the avoided cost is developed by dividing cost for a period by the consumption in the same period. This provides useful results for certain analyses, but such things as demand and fixed periodic charges may cause this approach to yield significant errors when calculating cost savings.

Simple facility layout maps should be provided to show building floorplans, define building areas, and illustrate major equipment locations. Additional layouts of convenient size (e.g., $8\frac{1}{2}'' \times 11''$) are useful for notations by assessment team members during both the walkthrough and the detailed assessment visit.

Walkthrough Assessments

Walkthrough assessments usually have three major outcomes: the identification of potential energy conservation projects; the identification of the effort, including skills, personnel and equipment required for a detailed assessment; and an estimation of the cost of a detailed assessment. The need for data recording may be identified at this time and arrangements to install sensors and loggers can begin. Self-contained, easily-installed loggers may be installed at this time (the walkthrough visit may be a useful time to place logging equipment and other requests for data if a detailed visit will follow at a reasonable time for retrieving the loggers and data).

The walkthrough visit also provides an opportunity to obtain plant information that will be valuable in conducting the detailed assessment. Such data includes building and industrial process line locations and functions, and the locations of major equipment, mechanical rooms, or electrical control centers, meters, and transformers. Major energy users and problems often can be identified during a walkthrough visit. A manager or senior employee familiar with processes, systems, and energy consumption should accompany the energy expert during the walkthrough assessment.

On occasion, the walkthrough assessment may reveal that further detailed assessment work would not be particularly fruitful, and thus it may end the assessment process. If a detailed assessment is to follow, the result of the walkthrough should be a letter-type report with a general identification of potential projects and an effort or cost estimate for the detailed assessment. Generally, there will be no calculation of savings in this initial report. The walkthrough (in addition to its preceding discussions) provides an opportunity to obtain information about any previous energy assessments. Plant personnel should also be asked about any in-house studies to save energy, reduce waste, and increase productivity that may have been undertaken.

Detailed Assessments

The goal of detailed assessments is to gather accurate technical data that will allow the assessment team to prepare a formal, technical report describing projects that can be implemented in the facility to contain costs. The effectiveness of cost saving projects generally depends on calculation of reductions in energy use, demand, and, if waste and productivity issues are also considered, then the cost effects of reducing waste or pollution and increasing productivity should also be calculated. Such calculations require carefully obtained, accurate data. The detailed assessment often involves a team of engineers and technicians, accompanied, when necessary, by facility personnel, compared to the walkthrough assessment which is usually performed by fewer persons.

Even for one-day visits such as those exemplified by DOE's IAC program, the visit can be divided into a tour of the plant equivalent to a walkthrough, and a period of detailed data gathering, equivalent to a detailed assessment, as personnel revisit key areas of the plant to gather data and observe processes.

At this time, data necessary to compose a good description of the facility should be collected if the preliminary activities and walkthrough have not satisfied this need. A good assessment report contains a facility description, complete with facility layout showing the location of major areas and equipment. The projects described in the report should refer to the applicable facility description section to maintain continuity and perspective.

The major data obtained, however, pertains to the cost containment projects to be described in the detailed assessment report. Energy projects fall into four major areas, the first three technical and the final one administrative in nature:

- Turning off equipment that is used or idling unnecessarily
- Replacing equipment with more efficient varieties
- Modifying equipment to operate more efficiently
- Administrative projects dealing with energy procurement

Often data for the administrative projects is obtained from contracts, billing histories, and interviews with managers and energy suppliers. Data for the administrative projects are relatively easier to obtain than that for the technical projects, and may in fact be obtained as part of preassessment or walkthrough activities.

Technical data for equipment turn-off, replacement, or modification generally is more difficult to obtain, and mostly will be obtained during the detailed assessment phase. It involves gathering equipment size data such as power or other capacity ratings, efficiencies, load factors (which may be a function of time), and operating hours. Size data is relatively easy to obtain from nameplate or other manufacturer specification information. Estimates of equipment sizes and efficiencies often are unreliable. Every major piece of equipment and system should be considered for possible savings, and equipment should be inspected for condition and to determine if it is operating properly. Efficiencies, while often not directly obtainable from data available in the facility, may be obtained from the manufacturer. Efficiency data as a function of load is desirable, because efficiency generally varies with load.

The most difficult and unreliable data to collect often are load factors and operating hours. Equipment loading is often estimated because measurements are expensive and difficult to make. Installing measurement equipment may require shutdown of the equipment being monitored, which may present production or operating problems. If the load varies with time, then measurements should be made for a long enough period to cover all possible loadings. If the load variation is predictable, covering one period is sufficient. If the variation is unpredictable, then the measurement only gives data useful to show what a turn-off, equipment change, or modification would have saved for that period. For equipment with a constant load, a one-time measurement will be adequate, but even a one-time measurement may be difficult to obtain. Because of the difficulty measuring and unpredictability of loading, load factors are often estimated. Savings are proportional to these load factors, and thus erroneous estimates of load factors can lead to large errors in savings estimates. Whenever possible, load factors should be based on measured data.

Operating hours can also be unreliable because they are often obtained from interviews with operators or supervisors, who may not understand exactly what is being asked, or who simply may not have accurate information. Time of operation errors are difficult to eliminate. Gross errors such as those that may be made about equipment turn-off after normal operating hours can sometimes be eliminated by an after-hours visit by the assessment team. Measurements of annual equipment operating hours are useful because project savings usually are measured on an annual basis. Only rarely is this possible (for example, a project subject to measurement and verification for performance contracting purposes might yield a year of measured data, but such projects are relatively rare). Relatively inexpensive measurements can be made of operating hours for short periods by self-contained loggers that measure equipment on/off times. The operating schedule for each building or production area is needed, including information about breaks in operation such as lunch and information about holidays.

Technical and physical data useful in identifying projects, calculating savings, designing conceptual projects or management procedures to capture the savings, estimating implementation costs, and composing the facility description may be obtained in any phase of the assessment process. Some information will have been obtained prior to the walkthrough assessment visit, but during the walkthrough, additional information can be obtained about operation, function, production, and building information. Function or production information for the various areas of the facility is needed. For all types of facilities, general construction information about the buildings should be gathered, including wall and roof types; wall and roof heights and lengths; and lighting and HVAC system types, numbers of units, controls, and operating hours. An inventory of equipment should also be assembled.

ASSESSMENT REPORTING

All energy assessment reports have a similar format. There should be an executive summary, a recommendation section, a facility description, and an energy consumption analysis, though not necessarily in that order. The recommendations are the heart of the report. Although they often appear after other major sections in older reports, modern usage has them appearing earlier, after the executive summary. The executive summary generally appears first, but is written last to summarize the other sections. Other parts such as disclaimers, acknowledgements, and appendices may be used as necessary.

Recommended Projects

Assessment recommendations are the most important part of the report and contain the technical analysis of the facility's energy usage. These recommendations should be clear, with the source of data documented and the analysis technique described with sample calculations. The details of manual calculations should be given and subject to independent verification. Analyses whose most important calculation is multiplying some portion of the utility bill by a "rule of thumb" percentage to establish savings for a project are not suitable.

Just as for energy analysis reports, individual assessment recommendations usually follow a fairly uniform outline. There is a short description followed by a summary of energy and cost savings, and of the implementation cost and financial measure of merit (often simple payback). Akin to the facility description section of an assessment report, there is an observation section which describes the existing situation, problems, and recommended changes, including designs or management techniques proposed to capture the savings. Then there is a calculation section which shows the results in energy, demand, and cost savings, and an implementation cost section, which develops the implementation cost and the financial measure of merit, either payback or some more sophisticated method.

If computer analysis is used to calculate savings, the program should be a recognized program applied by a trained user. Inputs and outputs should be carefully considered for accuracy and feasibility. If locally developed spreadsheets are used, then a sample calculation should be given in the body of the report for each important result. Avoid "black-box" analyses where little or nothing is known about the algorithm being applied in user-developed computer programs or spreadsheets that are not widely recognized.

For some assessment recommendations (e.g., a simple turn-off recommendation for equipment operating unnecessarily after hours), implementation can follow without further study. However, for more complex projects (e.g., replacing an air-blown material mover with a mechanical conveyor), additional design and engineering may be required before implementation. The design level of energy assessment reports is conceptual, not detailed, in nature. The cost analysis in the implementation section for such conceptual designs should give management sufficient information to decide whether to go forward with the project. Also, if additional design and engineering are required, then an investment in further study will be needed before a final decision.

When the result of an assessment recommendation is dependent upon the accomplishment of a separate recommendation, then the recommendation should be calculated in both the independent and dependent mode. For example, if an assessment recommendation reduces natural gas consumption in a facility, saving energy and money, and if in the same analysis a recommendation is made to change suppliers to achieve a less expensive cost of natural gas energy, then the project to change suppliers depends on the accomplishment of the project to reduce gas consumption; if it is not accomplished, the savings due to the supplier change will be greater. The effect of the project to change suppliers should be calculated both as though the consumption reduction would not occur (the independent case) and as if the consumption reduction would occur. In the latter, dependent case, the effect of changing suppliers is affected by the reduction in consumption. In this case, savings from the project to reduce consumption also depends on making the supplier change, and that project can be calculated in both the dependent and independent modes. To avoid a complicated mix of projects and calculations, in some cases where a single large project is composed of several smaller, related projects, a hierarchy of projects and dependencies can be established to guide the dependency calculations. The $100 million LoanSTAR program that placed energy-efficient retrofits in public buildings in the state of Texas pioneered this approach.[8]

Each project should refer to the plant description for the location of systems and for information about equipment; similarly, it should refer to the energy consumption section when that area is important to the recommendation.

Assessment recommendations are an acceptable place to show the environmental effect of each recommendation by calculating the carbon equivalent, or NO_x, reduction due to reducing energy use. Emission factors are available for various areas of the country that can be used for these calculations.

Implementation information, including warnings about possible personnel and equipment safety issues should be given in each assessment recommendation. Safety issues should receive prominent, obvious display.

Plant Description

Plant description information is obtained during all phases of the assessment process. This part of the report may be divided into a facility description and a process description, and it provides context for the rest of the report. The facility description describes the buildings, their construction, major production areas, and equipment locations. Transformers and meter locations for energy streams should be shown. A building layout should be included to show major production areas and equipment locations, particularly those important to the energy analysis being described in the report.

The process description should describe each major process that goes on in the plant, starting with the procurement of raw materials—giving sources and delivery methods—and proceeding until products are packaged, perhaps warehoused, and shipped. Emphasis should be placed on points in the process where major amounts of energy are consumed and on the machines that are involved in that energy usage. It is important to include a process flow diagram for each process.

At some point, either in the facility or process description, the major energy consuming equipment in the plant should be listed. If waste is to be considered, then points in the process where waste is generated are especially significant, and in that case a table of waste streams, handling equipment, and storage locations should be included.

Energy Consumption Data

Energy analysis data can be obtained for composing this important section of the report before the plant is visited. For an assessment that is broad in scope, the manufacturer should be asked to provide copies of a minimum of 12 consecutive months of energy bills for analysis. Sometimes even more data is desirable for a detailed audit. Shorter periods will not cover a complete annual weather cycle. Industrial analyses, unlike analyses of commercial and institutional buildings, usually are independent of weather. However, some industrial projects are weather dependent. Particularly in the case of industrial buildings which are fully climate controlled, weather may be an important factor, and thus bills representing the full 12 months of usage are an extremely helpful amount to seek. They do not need to be for an annual or particular fiscal period from the standpoint of technical analysis, but they should be consecutive and as recent as possible. Copies of actual bills should be sought because summaries likely will not give information such as the applicable tariff, any taxes, late fees, and demand and power factor data. For a broad-based assessment, bills for all important energy streams should be obtained. For special emphasis on equipment or areas of production, only bills for relevant energy streams should be considered.

From the bills, relevant tables and graphs of energy consumption, demand, power factor, and costs should be prepared for each account or meter. Careful review of this data often reveals important data about such things as demand control possibilities, equipment usage, and production changes. It provides a starting point for in-plant discussions with management about cost reduction possibilities. For plants that do not have sufficient personnel to review and consider tariffs and bills carefully, such problems as incorrect billing for state sales tax, late fees, inappropriate tariffs, and the occasional error in billings may be revealed to the energy auditor before the first plant visit.

At this stage, the avoided cost of energy (and demand, if applicable) for each energy stream can be determined. Avoided cost, sometimes incorrectly called marginal cost, is the amount that the plant will save (avoid paying) if energy use or demand is reduced by one unit. Electricity probably will have a demand component to its billing as well as an energy cost. Demand charges for natural gas [e.g., in \$/(MCF/day)/billing period], for steam [e.g., in \$/(1000 lb/day)/billing period], and possibly for other energy streams are infrequent. Common energy costs that are often used for rapid analyses and comparisons, such as the cost per unit of electrical energy obtained by dividing total electric cost in a period by the total electric energy consumed in kWh (a common method that blends energy, demand, and all other charges such as the customer or meter charge into the result) should be developed in this section.

This data can also be used to analyze electrical load factors of the types used by utilities (electric load factor is the consumption of energy in a billing period divided by the peak demand and the total number of hours in the billing period). The resulting values can be compared to nominal values. In addition, a load factor called the production load factor can be developed if the operating hours of the main production area of the plant are available (production load factor may be called operating load factor and is the consumption of energy in a billing period divided by the peak demand and the number of operating hours in the billing period). In this case, the computed load factor can be compared to unity, which would represent the best possible use of the plant's equipment during operating hours. Commonly, no plant or system actually achieves unity, so a production load factor of 75 or 85% may be considered good. As a diagnostic tool, if the production load factor exceeds unity, and nothing in the plant should be operating outside the main production hours, then unnecessary energy consumption after hours is indicated.

Executive Summary

This section of the assessment summarizes the report, and particularly the assessment recommendations. A table summarizing each assessment recommendation in the order that they occur in the report is very useful. Most often, these appear in descending order of estimated annual savings. The executive summary should give overall information about the total energy consumption and cost, as well as the common unit costs of energy that often are used for comparisons. Summary information about wastes and pollution prevention can be shown here as well.

The executive summary is the place to give and emphasize information that will enhance implementation of the recommended projects, and to provide warnings about safety and other implementation issues that may be repeated in the individual assessment recommendations.

CONCLUSION

The energy assessment process, including reporting, has been covered. For large facilities desiring to cut costs, the process that has been described will likely continue as a part of the cost-cutting process for a long time. Self-assessment procedures are available that can be used by smaller organizations that perhaps cannot afford to pay for an independent assessment.

These procedures include manuals to guide calculations[9] and a simple spreadsheet approach designed for use by facility operators untrained in energy assessments.[10] Benchmarking for commercial and institutional buildings is ahead of the industrial sector where important advances are still being made. There has been great diversity (in quality and meaning) of the data available to set standards of comparison for industry. The summer conference of the American Council for an Energy Efficient Economy in 2005 discussed industrial benchmarking[11] and more advances can be expected in this area.

REFERENCES

1. Capehart, B.L.; Spiller, M.B.; Frazier, S. Energy auditing. In *Energy Management Handbook*; Turner, W.C., Ed., Fairmont Press: Lilburn, GA, 2005; 23–39.
2. Energy Use and Management *2003 ASHRAE Handbook, Heating, Ventilating and Air-Conditioning Applications*; American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE): Atlanta, Georgia, 2003; 35.1–35.19.
3. http://www.oit.doe.gov/iac/ and http://iac.rutgers.edu/ (accessed August 2005).
4. http://www.eere.energy.gov/femp/ (accessed August 2005).
5. Energy Use and Management *2003 ASHRAE Handbook, Heating, Ventilating and Air-Conditioning Applications*; American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE): Atlanta, GA, 2003; 35.14.
6. Fuller, S.K.; Peterson, S.R. Life-cycle costing manual for the federal energy management program In *National Institute of Standard Technology Handbook*, US Government Printing Office: Washington, DC, 1996; Vol.135; http://www.bfrl.nist.gov/oae/publications/handbooks/135/pdf (accessed August 2005).
7. American National Standards Institute *A Management System for Energy*; ANSI/MSE 2000 Standard: New York, 2000.
8. *Texas LoanSTAR Program: Technical Guidelines and Format*, Texas State Energy Conservation Office: Austin, Texas, 1996, 40.
9. Muller, M.R.; Papadaratsakis, K. *Self-Assessment Workbook for Small Manufacturers*; Center for Advanced Energy Systems, Rutgers, the State University of New Jersey: Piscataway, NJ, 2003.
10. Ferland, K.; Kampschroer, N.; Heffington, W. M.; Huitink, D.; Kelley, R. *Texas Size Savings! A Step-by-Step Energy Assessment Guide and Calculator for Small Manufacturers*; Center for Energy and Environmental Resources, University of Texas at Austin, Austin, Texas, 2005.
11. 2005 ACEEE Summer Study on Energy Efficiency in Industry *Cutting the High Cost of Energy*; American Council for an Energy Efficient Economy: New York, 2005; (aceee_publications@aceee.org).

Auditing: Improved Accuracy

Barney L. Capehart
Department of Industrial and Systems Engineering, University of Florida College of Engineering, Gainesville, Florida, U.S.A.

Lynne C. Capehart
Consultant and Technical Writing Specialist, Gainesville, Florida, U.S.A.

Abstract
A frequent criticism of the quality and accuracy of energy audits is that they overestimate the savings potential for many customers. This entry discusses several problem areas that can potentially result in over-optimistic savings projections and suggests ways to increase quality and produce more accurate energy audits. Performing an energy and demand balance is the initial step a careful energy analyst should take when starting to evaluate the energy use at a facility. These balances allow one to determine what the largest energy users are in a facility, to find out whether all energy uses have been identified, and to check savings calculations by determining whether more savings have been identified than are actually achievable. Use of the average cost of electricity to calculate energy savings can give a false picture of the actual savings and may result in over-optimistic savings predictions. This entry discusses how to calculate the correct values from the electricity bills and when to use these values. Finally, the entry discusses several common energy-saving measures that are frequently recommended by energy auditors. Some of these may not actually save as much energy or demand as expected except in limited circumstances. Others have good energy-saving potential but they must be implemented carefully to avoid increasing energy use rather than decreasing it.

INTRODUCTION

Critics of energy audit recommendations often say that auditors overestimate the savings potential available to the customer. The possibility of overestimation concerns utilities who do not want to pay incentives for demand-side management programs if the facilities will not realize the expected results in energy or demand savings. Overestimates also make clients unhappy when their energy bills do not decrease as much as promised. The problem multiplies when a shared savings program is undertaken by the facility and an energy service company. Here, the difference between the audit projections and the actual metered and measured savings may be so significantly different that either there are no savings for the facility or the energy service company makes no profit.

More problems are likely to concern the accuracy of the energy audits for industrial manufacturing facilities and large buildings than for smaller commercial facilities because the equipment and operation of larger facilities is more complex. Based on our auditing experience over the last fifteen years, we have identified a number of areas where problems are likely to occur, and a number of these are presented and discussed. In addition, we have developed a few methods and approaches to dealing with these potential problems and we have found a few ways to initiate our energy audit analyses that lead us to more accurate results. One of these approaches is to collect data on the energy-using equipment in a facility and then to perform both an energy and a demand balance to help insure that we have reasonable estimates of energy uses—and therefore, energy savings—of this equipment.

Keywords: Audits; Energy audits; Energy analyses; Energy savings; Energy balances; Demand balances; Motor load factors.

CALCULATING ENERGY AND DEMAND BALANCES

The energy and demand balances for a facility are an accounting of the energy flows and the power used in the facility. These balances allow the energy analyst to track the energy and power inputs and outputs (uses) and to see if they match. A careful energy analyst should perform an energy and demand balance on a facility before developing and analyzing any energy management recommendations. This way, the analyst can determine what the largest energy users are in a facility, find out whether all or almost all energy uses have been identified, and see whether more savings have been identified than are actually achievable. Making energy use recommendations without utilizing the energy and demand balances is similar to making budget cutting recommendations without knowing exactly where the money is currently being spent.

When we perform an energy survey (audit), we inventory all of the major energy-using equipment in the facility. Then we list the equipment and estimate its energy consumption and demand using data gathered at the facility such as nameplate ratings of the equipment and operating hours. We develop our energy balance by major equipment categories such as lighting, motors, heating, ventilating and air conditioning (HVAC), air compressors, etc. We also have a category called miscellaneous to account for loads that we did not individually survey, such as copiers, electric typewriters, computers, and other plug loads. We typically allocate 10% of the actual energy use and demand to the miscellaneous category in the demand and energy balances. (For an office building instead of a manufacturing facility, this miscellaneous load might be 15%–20% because of extensive computer and other electronic plug loads.) Then we calculate the energy and demand for each of the other categories.

Lighting

The first major category we analyze is lighting because this is usually the category in which we have the most confidence in knowing the actual demand and hours of use. Thus, we believe that our energy and demand estimates for the lighting system are the most accurate, and can then be subtracted from the total actual use to let us continue to build up the energy and demand balance for the facility. We record the types of lamps and the number of lamps used in each area of the facility and ask the maintenance person to show us the replacement lamps and ballasts used. With this lamp and ballast wattage data together with a good estimate of the hours that the lights are on in the various areas, we can construct what we believe to be a fairly accurate description of the energy and demand for the lighting system. Operational hours of lighting are now easily obtained through the use of inexpensive minidata loggers. These minidata loggers are typically under $100 and can record detailed on-off data for a month or more at five minute intervals.

Air Conditioning

There is generally no other "easy" or "accurate" category to work on, so we proceed to either air conditioning or motors. In most facilities there will be some air conditioning, even if it is just for the offices that are usually part of the industrial or manufacturing facility. Many facilities—particularly here in the hot and humid southeast—are fully air conditioned. Electronics, printing, medical plastics and devices, and many assembly plants are common facilities that we see that are fully air conditioned. Boats, metal products, wood products, and plastic pipe manufacturing facilities are most often not air conditioned. Air conditioning system name plate data is usually available and readable on many units and efficiency ratings can be found from published Air-Conditioning and Refrigeration Institute (ARI) data[1] or from the manufacturers of the equipment. The biggest problem with air conditioning is to get runtime data that will allow us to determine the number of full-load equivalent operating hours for the air conditioning compressors or chillers. From our experience in north and north-central Florida, we use about 2200–2400 h per year of compressor runtime for facilities that have air conditioning which responds to outdoor temperature. Process cooling requirements are much different and typically have much larger numbers of full-load equivalent operating hours. With the equipment size, the efficiency data, and the full-load equivalent operating hours, we can construct a description of the energy and demand for the air conditioning system. Again, the minidata loggers can make a huge contribution to the accuracy of our audits by measuring the compressor and fan runtimes and by estimating the load factor on the compressor by measuring the line current into the compressor motor. As long as the load on the motor is 50% or above, this method gives a reasonably accurate result.

Motors

Turning next to motors, we begin looking at one of the most difficult categories to deal with in the absence of fully metered and measured load factors on each motor in the facility. In a one day plant visit, it is usually impossible to get actual data on the load factors for more than a few motors. Even then, that data is only good for the one day that it was taken. Very few energy auditing organizations can afford the time and effort to make long-term measurements of the load factor on each motor in a large facility. Thus, estimating motor load factors becomes a critical part of the energy and demand balance and also a critical part of the accuracy of the actual energy audit analysis. Using minidata loggers to estimate the motor load factor by measuring the line current is reasonably accurate as long as the motor is 50% loaded at least. Motor name plate data shows the horsepower rating, the manufacturer, and sometimes the efficiency. If not, the efficiency can usually be obtained from the manufacturer or from standard references such as the Energy-Efficient Motor Systems Handbook[2] or from software databases such as MotorMaster, produced by the Washington State Energy Office.[3] We inventory all motors over 1 hp and sometimes try to look at the smaller ones if we have time.

Motor runtime is another parameter that is very difficult to get, but the minidata loggers have solved this problem to a great extent. However, it is still not likely that a minidata logger will be placed on each motor over 1 hp in a large facility. When the motor is used in an application where it is constantly on, it is an easy case. Ventilating fans, circulating pumps, and some process drive motors are often in this class because they run for a known, constant period of time each year. In other cases, facility operating personnel must help provide estimates of motor runtimes.

With data on the horsepower, efficiency, load factor, and runtimes of motors we can construct a detailed table of motor energy and demands to use in our balances. Motor load factors will be discussed further in a later section.

Air Compressors

Air compressors are a special case of motor use with most of the same problems. Some help is available in this category because some air compressors have instruments showing the load factor and some have runtime indicators for hours of use. Otherwise, use of two minidata loggers on each air compressor is required. Most industrial and manufacturing facilities will have several air compressors, and this may lead to some questions as to which air compressors are actually used and for how many hours they are used. If the air compressors at a facility are priority scheduled, it may turn out that one or more of the compressors are operated continuously, and one or two smaller compressors are cycled or unloaded to modulate the need for compressed air. In this case, the load factors on the larger compressors may be unity. Using this data on the horsepower, efficiency, load factor, and runtimes of the compressors, we develop a detailed table of compressor energy use and demand for our energy and demand balances.

Other Process Equipment

Specialized process equipment must be analyzed on an individual basis because it will vary tremendously depending on the type of industry or manufacturing facility involved. Much of this equipment will utilize electric motors and will be covered in the motor category. Other electrically-powered equipment such as drying ovens, cooking ovens, welders, and laser and plasma cutters are nonmotor electric uses and must be treated separately. Equipment name plate ratings and hours of use are necessary to compute the energy and demand for these items. Process chillers are in another special class that is somewhat different from comfort air conditioning equipment because the operating hours and loads are driven by the process requirements and not the weather patterns and temperatures. Minidata loggers will be of significant help with many of these types of equipment.

Checking the Results

Once the complete energy and demand balances are constructed for the facility, we check to see if the cumulative energy/demand for these categories plus the miscellaneous category is substantially larger or smaller than the actual energy usage and demand over the year. If it is, and we are sure we have identified all of the major energy uses, we know that we have made a mistake somewhere in our assumptions. As mentioned above, one area that we have typically had difficulty with is the energy use of motors. Measuring the actual load factors is difficult on a one-day walkthrough audit visit, so we use our energy balance to help us estimate the likely load factors for the motors. We do this by adjusting the load factor estimates on a number of the motors to arrive at a satisfactory level of the energy and demand from the electric motors. Unless we do this, we are likely to overestimate the energy used by the motors, thus overestimating the energy savings from replacing standard motors with high-efficiency motors.

As an example, we performed an energy audit for one large manufacturing facility with a lot of motors. We first assumed that the load factors for the motors were approximately 80%, based on what the facility personnel told us. Using this load factor gave us a total energy use for the motors of over 16 million kWh/yr and a demand of over 2800 kW. Because the annual energy use for the entire facility was just over 11 million kWh/yr and the demand never exceeded 2250 kW, this load factor was clearly wrong. We adjusted the average motor load factor to 40% for most of the motors, which reduced our energy use to 9 million kWh and the demand to just under 1600 kW. These values are much more reasonable with motors making up a large part of the electrical load of this facility.

After we are satisfied with the energy/demand balances, we use a graphics program to draw a pie chart showing the distribution of energy/demand between the various categories. This allows us to visually represent which categories are responsible for the majority of the energy use. It also allows us to focus our energy-savings analyses on the areas of largest energy use.

PROBLEMS WITH ENERGY ANALYSIS CALCULATIONS

Over the course of performing over 200 large facility energy audits, we have identified a number of problem areas. One lies with the method of calculating energy cost savings (CS)—whether to use the average cost of electricity or break the cost down into energy and demand cost components. Other problems include instances where the energy and demand savings associated with specific energy efficiency measures may not be fully realized or where more research should go into determining the actual savings potential.

On-peak and Off-peak Uses: Overestimating Savings by Using the Average Cost of Electricity

One criticism of energy auditors is that they sometimes overestimate the dollar savings available from various energy efficiency measures. One way overestimation can result is when the analyst uses only the average cost of electricity to compute the savings. Because the average cost of electricity includes a demand component, using this average cost to compute the savings for companies who operate on more than one shift can overstate the dollar savings. This is because the energy cost during the off-peak hours does not include a demand charge. A fairly obvious example of this type of problem occurs when the average cost of electricity is used to calculate savings from installing high-efficiency security lighting. In this instance, there is no on-peak electricity use, but the savings will be calculated as if all the electricity was used on-peak.

The same problem arises when an energy efficiency measure does not result in an expected—or implicitly expected—demand reduction. Using a cost of electricity that includes demand in this instance will again overstate the dollar savings. Examples of energy efficiency measures that fall into this category are: occupancy sensors, photosensors, and adjustable speed drives (ASD). Although all of these measures can reduce the total amount of energy used by the equipment, there is no guarantee that the energy use will only occur during off-peak hours. While an occupancy sensor will save lighting kWh, it will not save any kW if the lights come on during the peak load period. Similarly, an ASD can save energy use for a motor, but if the motor needs its full load capability—as an air-conditioning fan motor or chilled water pump motor might—during the peak load period, the demand savings may not be there. The reduced use of the device or piece of equipment on peak load times may introduce a diversity factor that produces some demand savings. However, in most instances, even these savings will be overestimated by using the average cost of electricity.

On the other hand, some measures can be expected to provide their full demand savings at the time of the facility's peak load. Replacing 40 W T12 fluorescent lamps with 32 W super T8 lamps will provide a verifiable demand savings because the wattage reduction will be constant at all times and will specifically show up during the period of peak demand. Shifting loads to off-peak times should also produce verifiable demand savings. For example, putting a timer or energy management system control on a constant load electric drying oven to insure that it does not come on until the off-peak time will result in the full demand savings. Using high-efficiency motors also seems like it would produce verifiable savings because of its reduced kW load, but in some instances, there are other factors that tend to negate these benefits. This topic is discussed later on in this entry.

To help solve the problem of overestimating savings from using the average cost of electricity, we divide our energy savings calculations into demand savings and energy savings. In most instances, the energy savings for a particular piece of equipment is calculated by first determining the demand savings for that equipment and then multiplying by the total operating hours of the equipment. To calculate the annual CS, we use the following formula:

$$\begin{aligned}\text{CS} = {} & [\text{Demand Savings} \times \text{Average Monthly Demand} \\ & \text{Rate} \times 12\ \text{mo/yr}] + [\text{Energy Savings Average} \\ & \text{Cost of Electricity without Demand}]\end{aligned}$$

If a recommended measure has no demand savings, then the energy CS is simply the energy savings times the average cost of electricity without demand (or off-peak cost of electricity). This procedure forces us to think carefully about which equipment is used on-peak and which is used off-peak.

To demonstrate the difference in savings estimates, consider replacing a standard 30 hp motor with a high-efficiency motor. The efficiency of a standard 30 hp motor is 0.901 and a high-efficiency motor is 0.931. Assume the motor has a load factor of 40% and operates 8760 h/yr (three shifts). Assume also that the average cost of electricity is \$0.068/kWh, the average demand cost is \$3.79/kW/mo, and the average cost of electricity without demand is \$0.053/kWh. The equation for calculating the demand of a motor is:

$$D = \text{HP} \times \text{LF} \times 0.746 \times 1/\text{Eff}$$

The savings on demand (or demand reduction) from installing a high-efficiency motor is:

$$\begin{aligned}\text{DR} &= \text{HP} \times \text{LF} \times 0.746 \times (1/\text{Eff}_S - 1/\text{Eff}_H) \\ &= 30\ \text{hp} \times 0.40 \times 0.746\ \text{kW/hp} \times (1/0.901 - 1/0.931) \\ &= 0.32\ \text{kW}\end{aligned}$$

The annual energy savings are:

$$\text{ES} = \text{DR} \times \text{H} = 0.32\ \text{kW} \times 8760\ \text{h/yr} = 2803.2\ \text{kWh/yr}$$

Using the average cost of electricity above, the cost savings (CS_1) is calculated as

$$\begin{aligned}\text{CS}_1 &= \text{ES} \times (\text{Average cost of electricity}) \\ &= 2803.2\ \text{kWh/yr} \times \$0.068/\text{kWh} = \$190.62/\text{yr}\end{aligned}$$

Using the recommended formula above,

$$\begin{aligned}\text{CS} &= [\text{Demand Savings} \times \text{Average Monthly} \\ &\quad \text{Demand Rate} \times 12\ \text{mo/yr}] + [\text{Energy Savings} \times \\ &\quad \text{Average Cost of Electricity without Demand}] \\ &= (0.32\ \text{kW} \times \$3.79/\text{mo} \times 12\ \text{mo/yr}) + \\ &\quad (2803.2\ \text{kWh/yr} \times \$0.053/\text{kWh}) \\ &= (\$14.55 + \$148.57)/\text{yr} \\ &= \$163.12/\text{yr}\end{aligned}$$

In this example, using the average cost to calculate the energy CS overestimates the CS by \$27.50 per year, or 17%. Although the actual amount is small for one motor, if this error is repeated for all the motors for the entire facility as well as all other measures that only reduce the demand component during the off-peak hours, then the cumulative error in CS predictions can be substantial.

Motor Load Factors

Many of us in the energy auditing business started off assuming that motors ran at full load or near full load and based our energy consumption analysis and energy-saving analysis on that premise. Most books and publications that give a formula for finding the electrical load of a motor do not even include a term for the motor load factor. However, because experience showed us that few motors actually run at full load or near full load, we were left in a quandary about what load factor to actually use in our calculations because we rarely had good measurements on the actual motor load factor. A recent paper by R. Hoshide shed some light on the distribution of motor load factors provided from his experience.[4] In this paper, Hoshide noted that only about one-fourth of all three-phase motors run with a load factor greater than 60%, with 50% of all motors running at load factors between 30 and 60% and one-fourth running with load factors less than 30%. Thus, those of us who had been assuming that a typical motor load factor was around 70 or 80% had been greatly overestimating the savings from high-efficiency motors, ASD, high-efficiency belts, and other motor-related improvements.

The energy and demand balances discussed earlier also confirm that overall motor loads in most facilities cannot be anywhere near 70%–80%. Our experience in manufacturing facilities has been that motor load factors are more correctly identified as being in the 30%–40% range. With these load factors, we get very different savings estimates and economic results than when we assume that a motor is operating at a 70% or greater load factor as shown in our example earlier.

One place where the motor load factor is critical but often overlooked is in the savings calculations for ASD. Many motor and ASD manufacturers provide easy-to-use software that will determine savings with an ASD if you supply the load profile data. Usually, a sample profile is included that shows calculations for a motor operating at full load for some period of time and at a fairly high overall load factor—i.e., around 70%. If the motor only has a load factor of 50% or less to begin with, the savings estimates from a quick use of one of these programs may be greatly exaggerated. If you use the actual motor use profile with the load factor of 50%, you may find that the ASD will still save some energy and money, but often not as much as it looks like when the motor is assumed to run at the higher load factor. For example, a 20 hp motor may have been selected for use on a 15 hp load to insure that there is a "safety factor." Thus, the maximum load factor for the motor would be 75%. A typical fan or pump in an air conditioning system that is responding to outside weather conditions may only operate at its maximum load about 10% of the time. Because that maximum load here is only 15 hp, the average load factor for the motor might be more like 40%, and will not be even close to 75%.

High-Efficiency Motors

Another interesting problem area is associated with the use of high-efficiency motors. In Hoshide's paper mentioned earlier, he notes that in general, high-efficiency motors run at a faster full load speeds than standard efficiency motors.[4] This means that when a standard motor is replaced by a high-efficiency motor, the new motor will run somewhat faster than the old motor in almost every instance. This is a problem for motors that drive centrifugal fans and pumps because the higher operating speed means greater power use by the motor. Hoshide provides an example where he shows that a high-efficiency motor that should be saving about 5% energy and demand actually uses the same energy and demand as the old motor. This occurs because the increase in speed of the high-efficiency motor offsets the power savings by almost exactly the same 5% due to the cube law for centrifugal fans and pumps.

Few energy auditors ever monitor fans or pumps after replacing a standard motor with a high-efficiency motor; therefore, they have not realized that this effect has cancelled the expected energy and demand savings. Because Hoshide noted this feature of high-efficiency motors, we have been careful to make sure that our recommendations for replacing motors with centrifugal loads carry the notice that it will probably be necessary to adjust the drive pulleys or the drive system so that the load is operated at the same speed in order to achieve the expected savings.

Motor Belts and Drives

We have developed some significant questions about the use of cogged and synchronous belts and the associated estimates of energy savings. It seems fairly well accepted that cogged and synchronous belts do transmit more power dfrom a motor to a load than if standard smooth V-belts are used. In some instances, this should certainly result in some energy savings. A constant torque application like a conveyor drive may indeed save energy with a more efficient drive belt because the motor will be able to supply that torque with less effort. Consider also a feedback-controlled application such as a thermostatically-controlled ventilating fan or a level-controlled pump. In this case, the greater energy transmitted to the fan or pump should result in the task being accomplished faster than if less drive power was supplied and some energy savings should exist. However, if a fan or a pump operates in a nonfeedback application—as is very common for many motors—then there will not be any energy savings.

For example, a large ventilating fan that operates at full load continuously without any temperature or other feedback may not use less energy with an efficient drive belt because the fan may run faster as a result of the drive belt having less slip. Similarly, a pump which operates continuously to circulate water may not use less energy with an efficient drive belt. This is an area that needs some monitoring and metering studies to check the actual results.

Whether or not efficient drive belts result in any demand savings is another question. Because in many cases the motor is assumed to be supplying the same shaft horsepower with or without high-efficiency drive belts, a demand savings does not seem likely in these cases. Possibly using an efficient belt on a motor with a constant torque application that is controlled by an ASD might result in some demand savings. However, for the most common applications, the motor is still supplying the same load, and thus would have the same power demand. For feedback-controlled applications, there might be a diversity factor involved so that the reduced operation times could result in some demand savings, but not the full value otherwise expected. Thus, using average cost electricity to quantify the savings expected from high-efficiency drive belts could well overestimate the value of the savings. Verification of the cases where demand savings are to be expected is another area where more study and data are needed.

Adjustable Speed Drives (ASDs)

We would like to close this discussion with a return to ASDs because these are devices that offer a great potential for savings, but have far greater complexities than are often understood or appreciated. Fans and pumps form the largest class of applications where great energy savings are possible from the use of ASDs. This is a result again of the cube law for centrifugal fans and pumps—where the power required to drive a fan or pump is specified by the cube of the ratio of the flow rates involved. According to the cube law, a reduction in flow to one-half the original value could now be supplied by a motor using only one-eighth of the original horsepower. Thus, whenever an air flow or liquid flow can be reduced, such as in a variable air volume system or with a chilled water pump, there are dramatic savings possible with an ASD. In practice, there are two major problems with determining and achieving the expected savings.

The first problem is the one briefly mentioned earlier—determining the actual profile of the load involved. Simply using the standard profile in a piece of vendor's software is not likely to produce very realistic results. There are so many different conditions involved in fan and pump applications that taking actual measurements is the only way to get a very good idea of the savings that will occur with an ASD. Recent papers have discussed the problems with estimating the loads on fans and pumps and have shown how the cube law itself does not always give a reasonable value.[5–7] The Industrial Energy Center at Virginia Polytechnic Institute and Virginia Power Company have developed an approach where they classify potential ASD applications into eight different groups and then estimate the potential savings from analysis of each system and from measurements of that system's operation.[8] Using both an analytical approach and a few measurements allows them to get a reasonable estimate of the motor load profile and thus a reasonable estimate of the energy and demand savings possible.

The second problem is achieving the savings predicted for a particular fan or pump application. It is not enough just to identify the savings potential and then install an ASD on the fan or pump motor. In most applications, there is some kind of throttling or bypass action that results in almost the full horsepower still being required to drive the fan or pump most of the time. In these applications, the ASD will not save much unless the system is altered to remove the throttling or bypass device and a feedback sensor is installed to tell the ASD what fraction of its speed to deliver. This means that in many air flow systems, the dampers or vanes must be removed so that the quantity of air can be controlled by the ASD changing the speed of the fan motor. In addition, some kind of feedback sensor must be installed to measure the temperature or pressure in the system to send a signal to the ASD or a programmable logic controller (PLC) to alter the speed of the motor to meet the desired condition. The additional cost of the needed alterations to the system and the cost of the control system greatly change the economics of an ASD application compared to the case where only the purchase cost and installation cost of the actual ASD unit is considered.

For example, a dust collector system might originally be operating with a large 150 hp fan motor running continuously to pick up the dust from eight saws. However, because production follows existing orders for the product, sometimes only two, three, or four saws are in operation at a particular time. Thus, the load on the dust collector is much lower at these times than if all eight saws were in use. An ASD is a common recommendation in this case, but estimating the savings is not easy to begin with and once the costs of altering the collection duct system and the cost of adding a sophisticated control system to the ASD is considered, the bottom line result is much different than the cost of the basic ASD with installation. Manual or automatic dampers must be added to each duct at a saw so that it can be shut off when the saw is not running. In addition, a PLC for the ASD must be added to the new system together with sensors added to each damper so that the PLC will know how many saws are in operation and therefore what speed to tell the ASD for the fan to run to meet the dust collection load of that number of saws. Without these system changes and control additions, the

ASD itself will not save any great amount of energy or money. Adding them in might double the cost of the basic ASD and double the payback time that may have originally been envisioned.

Similarly, for a water or other liquid flow application, the system piping or valving must be altered to remove any throttling or bypass valves and a feedback sensor must be installed to allow the ASD to know what speed to operate the pump motor. If several sensors are involved in the application, a PLC may also be needed to control the ASD. For example, putting an ASD on a chilled water pump for a facility is much more involved and much more costly than simply cutting the electric supply lines to the pump motor and inserting an ASD for the motor. Without the system alterations and without the feedback control system, the ASD cannot provide the savings expected.

CONCLUSION

Energy auditing is not an exact science, but a number of opportunities are available for improving the accuracy and the quality of the recommendations. Techniques that may be appropriate for small-scale energy audits can introduce significant errors into the analyses for large complex facilities. We began by discussing how to perform an energy and demand balance for a company. This balance is an important step in doing an energy-use analysis because it provides a check on the accuracy of some of the assumptions necessary to calculate savings potential. We also addressed several problem areas that can result in over-optimistic savings projections and suggested ways to prevent mistakes. Finally, several areas where additional research, analysis, and data collection are needed were identified. Once this additional information is obtained, we can produce better and more accurate energy audit results.[9]

REFERENCES

1. Air-Conditioning and Refrigeration Institute, *ARI Unitary Directory*, published every six months. Also available through www.ARI.org.
2. Nadel, S.; Shepard, M.; Greenberg, S.; Katz, G.; de Almeida, A. *Energy-Efficient Motor Systems: A Handbook on Technology, Program and Policy Opportunities*, 2nd Ed.; American Council for an Energy-Efficient Economy: Washington, DC, 2002.
3. Washington State Energy Office, MotorMaster Electric Motor Selection Software and Database, updated annually. Available from the U.S. Department of Energy, at www.eere.energy.gov/industry 2006.
4. Hoshide, R.K. Electric Motor Do's and Don'ts. Energy Eng. **1994**, *91* (1).
5. Stebbins, W.L. Are you certain you understand the economics for applying ASD systems to centrifugal loads? Energy Eng. **1994**, *91* (1).
6. Vaillencourt, R.R. Simple solutions to VSD pumping measures. Energy Eng. **1994**, *91* (1).
7. Kempers, G. DSM pitfalls for centrifugal pumps and fans. Energy Eng. **1995**, *92* (2).
8. Personal communication with Mr. Mark Webb, Senior Engineer, Virginia Power Company: Roanoke, VA, January 1995.
9. Capehart, B.; Capehart, L. Improving industrial energy audit analysis. In *Proceedings of the 1995 ACEEE Summer Study in Industry*, ACEEE: Washington, DC, August 2005.

BIBLIOGRAPHY

1. Capehart, B.L.; Turner, W.C.; Kennedy, W.G. *Guide to Energy Management*, 5th Ed.; Fairmont Press: Lilburn, GA, 2005.
2. Turner, W.C., Ed. *Energy Management Handbook* 5th Ed., Fairmont Press: Lilburn, GA, 2006.
3. Thumann, A.; Younger, W.J. *Handbook of Energy Audits*, 6th Ed.; Fairmont Press: Lilburn, GA, 2003.
4. Thumann, A. *Plant Engineers and Managers Guide to Energy Conservation*, 7th Ed.; Fairmont Press: Lilburn, GA, 2006.

Auditing: User-Friendly Reports

Lynne C. Capehart
Consultant and Technical Writing Specialist, Gainesville, Florida, U.S.A.

Abstract
Energy audits do not save money and energy for companies unless the recommendations are implemented. Audit reports should be designed to encourage implementation, but often they impede it instead. In this article, the author discusses her experience with writing industrial energy audit reports and suggests some ways to make the reports more user-friendly. The goal in writing an audit report should not be the report itself; rather, it should be to achieve implementation of the report recommendations and thus achieve increased energy efficiency and energy cost savings for the customer.

INTRODUCTION

This article addresses two questions: "Why should an energy audit report be user-friendly?" and "How do you make an audit report user-friendly?" The author answers these questions in the context of sharing experience gained by writing audit reports for industrial clients of the University of Florida Industrial Assessment Center (UF IAC).

At the UF IAC, we had two goals in writing an audit report. Our first goal was to provide our clients with the facts necessary to make informed decisions about our report recommendations. Our second goal, which was as important as the first, was to interest our clients in implementing as many of our recommendations as possible. We found that "user-friendly" audit reports helped us achieve both goals.

WHAT IS A USER-FRIENDLY AUDIT REPORT?

The definition of "user-friendly" is something that is easy to learn or to use. People generally think of the term "user-friendly" related to something like a computer program. A program that is user-friendly is one that you can use with minimal difficulty. We have applied the same term to audit reports to mean a report that communicates its information to the user (reader) with a minimum amount of effort on the reader's part. We operate on the belief that a reader who is busy will not want to spend valuable time struggling to understand what the report is trying to say. If the report is not clear and easy to follow, the reader will probably set it down to read later, and "later" may never come!

Keywords: Audit reports; Energy auditing; Audit recommendations; Report writing; Technical writing.

HOW DO YOU WRITE A USER-FRIENDLY AUDIT REPORT?

From our experience, we have identified a number of key points for successfully writing a user-friendly audit report. These points are summarized below.

Know your audience

The first thing to keep in mind when you start to write anything is to know who your audience is and to tailor your writing to that audience. When writing an industrial audit report, your readers can range from the company president to the head of maintenance. If recommendations affect a number of groups in the company, each group leader may be given a copy of the report. Thus, you may have persons of varying backgrounds and degrees of education reading the report. Not all of them will necessarily have a technical background. The primary decision maker may not be an engineer; the person who implements the recommendations may not have a college degree.

We dealt with this problem by writing a report with three basic sections. Section One was an executive summary that briefly described our recommendations and tabulated our results such as the energy and dollar savings and the simple payback times. Section Two was a brief description of a recommended energy management program for our client. Section Three was a detailed section that we called our technical supplement. This section of our report included the calculations that supported our recommendations and any specific information relating to implementation. (These sections are described more fully later in this article.)

Use a simple, direct writing style

Technical writers often feel compelled to write in a third-person, passive, verbose style. Because energy audit reports are technical in nature, they often reflect this writing style. Instead, you should write your audit report in clear, understandable language. As noted above, your reader may not have a technical background. Even a reader who does will not be offended if the report is easy to read and understand. Some specific suggestions are:

Simplify your writing by using active voice. Technical writers use passive voice, saying "It is recommended..." or "It has been shown..." rather than "We recommend..." or "We have shown..." Passive voice allows the writer to avoid taking direct responsibility for the recommendations. Be clear and straightforward in your writing by using active voice wherever possible.

Address the report to the reader. Write as if you were speaking directly to the reader. Use the words "you" and "your." Say "your company…," "your electric bill…," etc. Make the report plain and simple. The following examples show how to do this.

Not: Installation of high-efficiency fluorescent lamps in place of the present lamps is recommended.
But: Install high-efficiency fluorescent lamps in place of your present lamps.
Or: We recommend that you install high-efficiency fluorescent lamps in place of your present lamps.
Not: Twelve air leaks were found in the compressor system during the audit of this facility.
But: We found twelve air leaks in the compressor system when we audited your facility.
Or: You have twelve air leaks in your compressor system.

Avoid technical jargon that your reader may not understand. Do not use acronyms such as ECO, EMO, or EMR without explaining them. (Energy Conservation Opportunity, Energy Management Opportunity, Energy Management Recommendation).

Present Information Visually

Often the concepts you need to convey in an audit report are not easy to explain in a limited number of words. To solve this problem, we often used drawings to show what we meant. For example, we had a diagram that showed how to place the lamps in fluorescent lighting fixtures when you are eliminating two of the lamps in a four-lamp fixture and adding reflectors. We also had a diagram showing how a heat pipe works.

We also presented our client's energy use data visually with graphs showing the annual energy and demand usage by month. These graphs gave a picture of use patterns. Any discrepancies in use showed up clearly.

Make Calculation Sections Helpful

The methodology and calculations used to develop specific energy management opportunity recommendations can be very helpful in an audit report. When you include the methodology and calculations, the technical personnel have the opportunity to check the accuracy of your assumptions and your work. Because not every reader wants to wade through pages describing the methodology and showing the calculations, we provided this information in a technical supplement to our audit report. Because this section was clearly labeled as the technical supplement, nontechnical readers could see instantly that this section might be difficult for them to understand and that they could ignore it.

Use Commonly Understood Units

In your report, be sure to use units that your client will understand. Discussing energy savings in terms of BTUs is not meaningful to the average reader. Kilowatt-hours for electricity or therms for natural gas are better units because most energy bills use these units.

Make Your Recommendations Clear

Some writers assume that their readers will understand their recommendation even if it is not explicitly stated. Although the implication may often be clear, better practice is to clearly state your recommendation so that your reader knows exactly what to do.

Not: Install occupancy sensors in the conference room and restrooms.
But: You should purchase five occupancy sensors. Install one in the conference room and one in each of the four restrooms.

Explain Your Assumptions

A major problem with many reports is the author's failure to explain the assumptions underlying the calculations. For example, when you use operating hours in a calculation, show how you got the number. "Your facility operates from 7:30 A.M. to 8:00 P.M., 5 days a week, 51 weeks per year. Therefore, we will use 3188 annual operating hours in our calculations."

When you show your basic assumptions and calculations, the reader can make adjustments if those facts change. In the example above, if the facility decided to operate 24 h/day, the reader would know where and how to make changes in operating hours because we had clearly labeled that calculation.

Use a section of your report to list your standard assumptions and calculations. That way, you do not have to repeat the explanations for each of your recommendations. Some of the standard assumptions/calculations that can be included in this section are operating hours, the average cost of electricity, the demand rate, the off-peak cost of electricity, and the calculation of the fraction of air conditioning load attributable to lighting.

Be Accurate and Consistent

The integrity of a report is grounded in its accuracy. This does not just pertain to the correctness of calculations. Clearly, inaccurate calculations will destroy a report's credibility, but other problems can also undermine the value of your report.

Be consistent throughout the report. Use the same terminology so your reader is not confused. Make sure that you use the same values. Do not use two different load factors for the same piece of equipment in different recommendations. For example, you might calculate the loss of energy due to leaks from an air compressor in one recommendation and the energy savings due to replacing the air compressor motor with a high efficiency motor in another recommendation. If you use different load factors or different motor efficiencies in each recommendation, your results will not be consistent or accurate.

Proofread your report carefully. Typographical and spelling errors devalue an otherwise good product. With computer spell checkers, there is very little excuse for misspelled words. Your nontechnical readers are likely to notice this type of error, and they will wonder if your technical calculations are similarly flawed. Textual errors can also sometimes change the meaning of a sentence—if you say "Do not…" instead of "Do…," you have made a major mistake.

REPORT SECTIONS

We found that the following report format met our clients' needs and fit our definition of a user-friendly report.

Executive Summary

The audit report starts with an executive summary, which lists the recommended energy conservation measures and shows the implementation cost and dollar savings amount. This section is intended for the readers who only want to see the bottom line. Although the executive summary can be as simple as a short table, you may add brief text to explain the recommendations and include any special information needed to implement the recommendations. We copied the executive summary on colored paper so that it stood out from the rest of the report.

Energy Management Plan

Following the executive summary, we provided some information to the decision makers on how to set up an energy management program in their facility. We viewed this section as one that encouraged implementation of our report, so we tried to make it as helpful as possible.

Energy Action Plan. In this subsection, we described the steps that a company should consider in order to start implementing our recommendations.

Energy Financing Options. We also included a short discussion of the ways that a company can pay for the recommendations. We covered the traditional use of company capital, loans for small businesses, utility incentive programs, and the shared savings approach of the energy service companies.

Maintenance Recommendations. We did not usually make formal maintenance recommendations in the technical supplement because the savings are not often easy to quantify. However, in this section of the report, we provided energy-savings maintenance checklists for lighting, heating/ventilation/air-conditioning, and boilers.

The Technical Supplement

The technical supplement is the part of the report that contains the specific information about the facility and the audit recommendations. Our technical supplement had two main sections: one included our assumptions and general calculations and the other described the recommendations in detail, including the calculations and methodology. We sometimes included a third section that described measures we had analyzed and determined were not cost-effective or that had payback times beyond the client's current planning horizon.

Standard Calculations and Assumptions

This section was briefly described above when we discussed the importance of explaining assumptions. Here we provided the reader with the basis for understanding many of our calculations and assumptions. We included a short description of the facility: square footage (both air conditioned and unconditioned areas), materials of construction, type and level of insulation, etc. If we were dividing the facility into subareas, we described those areas and assigned each an area number that was then used throughout the recommendation section.

Standard values calculated in this section included operating hours, the average cost of electricity, the demand rate, the off-peak cost of electricity, and the calculation of the fraction of air conditioning load attributable to lighting. When we calculated a value in this section, we labeled the variable with an identifier that remained the same throughout the rest of the report. For example, operating hours was OH wherever it was used; demand rate was DR.

Audit Recommendations

This section contained a discussion of each of the energy management opportunities we had determined were cost-effective. Each energy management recommendation (or EMR) that was capsulized in the executive summary was described in depth here.

Again, we tried to make the EMRs user-friendly. To do this, we put the narrative discussion at the beginning of the recommendation and left the technical calculations for the very end. This way, we allowed the readers to decide for themselves whether they wanted to wade through the specific calculations.

Each EMR started with a table that summarized the energy, demand, and cost savings, implementation cost, and simple payback period. Then we wrote a short narrative section that provided some brief background information about the recommended measure and explained how it should be implemented at this facility. If we were recommending installation of more than one item (lights, motors, air conditioning units, etc.), we often used a table to break down the savings by unit or by area.

The final section of each EMR was the calculation section. Here we explained the methodology that we used to arrive at our savings estimates. We provided the equations and showed how the calculations were performed so that our clients could see what we had done. If they wanted to change our assumptions, they could. If some of the data we had used was incorrect, they could replace it with the correct data and recalculate the results. However, by placing the calculations away from the rest of the discussion rather than intermingling the two, we didn't scare off the readers who needed to look at the other information.

Appendix

We used an appendix for lengthy data tables. For example, we had a motor efficiencies table that we used in several of our EMRs. Instead of repeating it in each EMR, we put it in the appendix. We also included a table showing the facility's monthly energy-use history and a table listing the major energy-using equipment. Similar to the calculation section of the EMRs, the appendix allowed us to provide backup information without cluttering up the main body of the report.

SHORT FORM AUDIT REPORT

Many energy auditors use a short form audit report. A short report is essential when the cost of the audit is a factor. Writing a long report can be time-consuming and it increases the cost of an audit.

The short form report is useful when an on-the-spot audit report is required because the auditor can use a laptop computer to generate it. It is also an excellent format for preliminary audit reports when the company will have to do further analysis before implementing most of the recommendations.

However, some short form audit reports have drawbacks. When a report is ultra-short and only provides the basic numbers, the reader will not have a memory crutch if he returns to the report sometime after the auditor has left. Because some clients do not implement the recommendations immediately but wait until they gather the necessary capital, an ultra-short form report may lose its value. Therefore, some explanatory text is a critical component of a user-friendly short form report. The executive summary described above could serve as a model short form audit report.

FEEDBACK

Customer feedback is as appropriate in energy auditing as in any other endeavor. An easy way to get feedback is to give the customer a questionnaire to evaluate the audit service and the report. In our feedback form, we listed each section of the report and asked the client to rate each section on a scale of 1–10, with 1 being poor and 10 being excellent. We asked for a rating based on whether the section was easy to read and we asked for a second rating of the likelihood that our recommendations would be implemented. (We also asked for any additional comments, but seldom got those.)

The questionnaire must be easy to fill out. If it takes too much time to read and fill out, the clients won't take the time to return it. We used to send the questionnaire along with the report, but those were seldom returned. We decided to wait for a month and then send the questionnaire as a follow-up to the audit. We had a much greater return rate when we used this method.

CONCLUSION

Many audit reports are not user-friendly. Most often they are either lengthy documents full of explanations, justifications, and calculations or they are very short with little backup information. If a report is so long that it intimidates your readers by its very size, they may set it aside to read when they have more time. If it is so short that needed information is lacking, the readers may not believe the results.

Writing a user-friendly audit report is an important step in promoting implementation of audit recommendations. If you adopt some of these report-writing suggestions, you should be able to produce your own successful user-friendly energy audit report.

ACKNOWLEDGMENT

An earlier version of this article appeared in Strategic Planning for Energy and the Environment, and it is used with permission from Fairmont Press.

BIBLIOGRAPHY

1. How to Write in Plain English, http://www.plainenglish.co.uk/plainenglishguide.html.
2. A Short Course on Writing Technical Reports, http://www.technical-writing-course.com/type-of-technical-report.html.
3. Technical Writing and Writing a Technical Report, www.engr.arizona.edu/robotics/Robotics/Technical%20Report%20Guidelines.doc.
4. Short Reports: How to Write Routine Technical Documents, http://jerz.setonhill.edu/writing/technical/reports/index.html.
5. Strunk, W. Jr; White, E.B.; Angell, R. *The Elements of Style*, 4th Ed.; 2000.
6. Philip, R. *Science and Technical Writing: A Manual of Style*, 2nd Ed.; Routledge Study Guides S.: London, 2000.

Benefit Cost Analysis

Fatouh Al-Ragom
Building and Energy Technologies Department, Kuwait Institute for Scientific Research, Safat, Kuwait

Abstract
Benefit cost analysis is one of several methods utilized to evaluate the feasibility of capital investment. The benefit cost analysis calculates the present worth of all benefits, then calculates the present worth of all costs and takes the ratio of the two sums. This ratio is either known as a benefit/cost ratio, a savings/interest ratio, or a profitability index. The benefit cost analysis is an economic decision-making criterion that can measure the economic consequences of a decision over a specified period of time. It is used to evaluate whether an alternative option is economically viable when compared to a base case which is usually the "do nothing" option or it can be used to rank several options that are competing for a limited budget.

INTRODUCTION

The implementation of energy efficiency projects is linked to the allocation of funds. Economic decision-making tools or methods are often required to justify a project's implementation and assess its economical feasibility. Several methods are available, such as life-cycle cost, simple or discounted payback, benefit cost analysis, and internal and adjusted rate-of-return.[1]

The selection of the method depends on many factors and, in fact, more than one method can be technically appropriate for economical decisions. These methods are often utilized to compare several options or to compare an alternative to the current situation usually referred to as the "do nothing" option. The American Society for Testing and Materials (ASTM) provides a guide that details the selection of economic methods for evaluating investments.[2]

This entry describes the economic assessment method known as benefit cost analysis (BCA). This method is simply considered as an attempt to identify and express, in dollar terms, all of the effects of proposed policies or projects and then to relate those effects with endured costs in a simple, dimensionless parameter.

BACKGROUND

Benefit cost analysis is one of the economic assessment methods that can aid policymakers in deciding whether the advantages of a particular course of action are likely to outweigh its drawbacks.

Keywords: Economical assessment; Decision-making; Resource-allocation decisions; Benefit-to-cost-ratio; Savings-to-investment-ratio; Profitability index.

Benefit cost analysis was conceived over 150 years ago by the French engineer Jules Dupuit. The BCA method saw its first widespread use when the United States government wanted to evaluate the impact of water projects in the late 1930s. Since then, it has been used to analyze policies affecting several sectors including public health, transportation, criminal justice, education, defense, and the environment.[7]

The BCA method involves evaluating benefits of a project against its costs in a ratio format that is known as benefit-to-cost-ratio (BCR). Another variation of BCR is known as savings-to-investment-ratio (SIR). The difference between the two is that BCR is used when the focus of the analysis is on benefits (that is, advantages measured in dollars) relative to project costs while SIR is used when the focus of the assessment is on project savings (that is, cost reductions).[1] The BCR is referred to as the profitability index (PI) in some financial publications.[3] The estimation of a project's PI is usually carried out based on the organization's current net value of cash flows.

SIGNIFICANCE OF BCR

The BCR provides a standard for measuring economic performance in a single number that indicates whether a proposed project or system is preferred over a mutually exclusive alternative that serves as the baseline for the economic analysis. In addition, the BCR indicates the discounted dollar benefits (or savings) per dollar of discounted costs. Moreover, it can be used to determine if a given project or system is economically feasible relative to the alternative of not implementing it. Also, when computed on increments of benefits (or savings) and costs, the BCR can be used to determine if one design or size of a system is more economic than another. On the other hand, when funding is limited, the BCR can be used as an aid to select the most economically efficient set of projects from among several available options that are competing for limited funding. Selecting an efficient set of projects will maximize aggregate net benefits or net savings obtainable for the budget.[1] The BCR and the PI examine cash flows, not accounting profits, and recognize the time-value of money. While these ratios can be accurate predictors of economic efficiency, their accuracy depends on the accuracy of cash flow predictions.

PROCEDURE

Before conducting a BCA, multiple implementation alternatives should be identified in a way that allows for a fair comparison. The constraints or requirements of a successful end solution should be also clearly identified. Additionally, costs and benefits should be put into standard units (usually dollars) so that they can be compared directly. The dollar amounts used in the BCA should all be discounted, that is, expressed in time-equivalent dollars, either in present value or uniform annual value terms.

The BCR is a numerical ratio that indicates the expected economic performance of a project by the size of the ratio. A ratio less than 1.0 indicates a project that is uneconomic, a ratio of 1.0 indicates a project whose benefits or savings are just equal to its costs, and a ratio greater than 1.0 indicates a project that is economic.

The recommended steps for carrying out an economic evaluation using the BCA method are summarized as follows:

- *Identification of objectives, constraints, and alternatives*. The decision-maker's objectives should be clearly specified. This is crucial to defining the problem. Moreover, constraints that limit potential alternatives for accomplishing the objectives should be identified such as economic and environmental limitations. Finally, alternatives that are technically and otherwise feasible in view of the constraints should be identified for consideration.
- *Data collection*. Actual or expected cash flows are needed for all phases of the project including revenues or other benefits; acquisition costs, including costs of planning, design, construction, purchase, installation, site preparation; utility costs, including costs of energy and operating and maintenance costs; repair and replacement costs; salvage values; disposal costs; and insurance costs. Moreover, information is also needed regarding the study period, discount rate, any applicable tax rates and rules, and the terms of financing (if applicable).
- *Expression of cash flow*. A decision should be made about whether to express the discounted cash flow of costs and benefits within each year in present-value dollars or in annual-value dollars. This should also include deciding whether to work in constant dollars using a real discount rate or in current dollars using a nominal discount rate. When using constant dollars, inflation is not included in the estimates of costs and benefits.[5]
- *Compute the BCR or SIR*. In concept, the BCA ratios are simple: benefits (or savings) divided by costs, where all dollar amounts are discounted to current or annual values. The BCR will be computed using the following formula:

$$\mathrm{BCR} = \frac{\sum_{t=0}^{n} B_t/(1+i)^t}{\sum_{t=0}^{n} C_t/(1+i)^t} \tag{1}$$

Where B_t, benefits in period t; C_t, costs in period t; n, life of the investment; i, discount rate

Note that $1/(1+i)^t$ is the discount factor used to calculate the present worth of a single future payment. Thus, to account for all future values, each value is calculated separately and then all values are added together. When evaluating an energy saving opportunity, its benefit at the beginning (at $t=0$) will be zero, while for the following time periods the benefits (savings) will be evaluated separately. Usually, the costs will only be considered at $t=0$ representing the initial cost. If the analysis will be carried out utilizing the uniform series of present worth as in the case of constant annual energy savings, then the following equation can be utilized to estimate the BCR:

$$\mathrm{BCR} = \frac{B((1+i)^n - 1)/i(1+i)^n}{C} = \frac{B \times (P/A,i\%,n)}{C} \tag{2}$$

Where B, annual benefits; C, initial cost; $(P/A, i\%, n)$, discount factor for a uniform series of cash flows

It should be noted that the BCR and PI represent the same parameter; and thus the PI can be calculated using Eq. 1. Alternatively Eq. 3 can be also used.

$$\mathrm{PI} = \frac{\mathrm{PV\ (future\ cash\ flows)}}{\mathrm{initial\ investment}} = 1 + \frac{\mathrm{NPV}}{\mathrm{initial\ investement}} \tag{3}$$

Where PV, Present value; NPV, Net present value of cash = benefits (inflows) − expenses (outflows)

Example 1

A building owner has \$12,000 available to utilize for improvements. The owner is considering replacing the old

Table 1 Cash flow details

n (Year)	(1): Annual savings	(2): $1/(1+i)^n$	PV=(1)×(2)
0	0	1	0
1	1765	0.952	1681
2	1765	0.907	1600.9
3	1765	0.864	1524.7
4	1765	0.823	1452.1
5	1765	0.784	1382.9
6	1765	0.746	1317.1
7	1765	0.711	1254.4
8	1765	0.677	1194.6
9	1765	0.645	1137.7
10	1765	0.614	1083.6
			13,629

boiler that has an efficiency of 65% with a new one that has an efficiency of 85%. Assuming that the lifespan of the boiler is 10 years and the discount rate is 5%, evaluate if this investment is cost-effective or not. The boiler uses oil and consumes 6000 gallons/year at a cost of \$1.25 per gallon.

This example compares the option of using a new boiler to the "do nothing" option of using the old boiler. Accordingly, due to the higher efficiency of the new boiler, annual operational costs will be lower, assuming that the salvage values of the old boiler and the new one are insignificant.

The annual savings due to the new boiler's efficiency improvement are calculated using the following equation[4]:

$$\text{Fuel Savings} = \frac{\eta_{\text{New}} - \eta_{\text{Old}}}{\eta_{\text{New}}} \times \text{Old Fuel Consumption} \times \text{Fuel Cost} \quad (4)$$

$$\text{Fuel Savings} = \frac{0.85 - 0.65}{0.85} \times 6000 \text{ gallons/year} \times 1.25/\text{gallons} = 1765$$

The cash flows are detailed in Table 1, and are calculated year by year.

An alternative option to estimating the present worth of future savings would be to use a uniform series present worth factor. This is done for ease of calculation. Instead of discounting the amounts in each year and adding them together, the cash flows can be grouped into categories with the same pattern of occurrence and discounted using discount factors. In this case, since there is a uniform savings of \$1,765 per year, the present value of the future annual cash flows is calculated using the following equation[8]:

$$\text{PV} = A\,\frac{(1+i)^n - 1}{i(1+i)^n} = A \times (P/A,i\%,n) \quad (5)$$

$$\text{PV} = 1765 \times (P/A,5\%,10) = 1765 \times 7.722 = 13{,}629$$

Where PV, present value of future annual cash flows; A, annual savings; $(P/A,\ i\%,\ n)$, discount factor for uniform series of cash flows. The values of this factor are listed in financial analysis books. In this example, the value of $(P/A,\ 5\%,\ 10) = 7.722$.[6]

The BCR using Eq. 2 will be calculated as:

$$\text{BCR} = \frac{\$13{,}629}{\$12{,}000} = 1.136$$

Since the BCR is more than one, this means that the benefits outweigh the costs; hence, the boiler replacement is financially justified.

- In practice, it is important to formulate the ratio so as to satisfy the investor's objective. This requires attention to the placement of costs in the numerator and denominator. To maximize net benefits from a designated expenditure, it is necessary to place only that portion of costs in the denominator on which the investor wishes to maximize returns. Then, the BCR will be computed using the following formula:

$$\text{BCR} = \frac{\sum_{t=0}^{n}(B_t - C_t)/(1+i)^t}{\sum_{t=0}^{n}\text{IC}_t/(1+i)^t} \quad (6)$$

  Where B_t, benefits in period t; that is, the advantages in revenue or performance (measured in dollars) of the building or system as compared to a mutually exclusive alternative; C_t, costs in period t, excluding investment costs that are to be placed in the denominator for the building or system, minus counterpart costs in period t for a mutually exclusive alternative; IC_t, investment costs in period t on which the investor wishes to maximize the return, minus similar investment costs in period t for a mutually exclusive alternative; i, discount rate.
- Special attention should be paid when placing cost and benefit items into the ratio especially in the case when several projects are compared and a ranking of their cost-effectiveness needs to be established. Changing the placement of a cost item from the denominator (where it increases costs) to the numerator (where it

Table 2 Investment details among alternative projects

Project code	(1) Project description	(2) Project cost PV ($)	(3) Energy savings PV ($)	(4) BCR	(5) Ranking	(6) Net benefits	(7) PI
A	Wall R-10	6,050	12,000	1.983	2	5950	1.983
B	Wall R-15	6,900	13,800	2.000	1	6900	2.000
C	Roof R-15	7,200	5,000	0.694	5	−2200	0.694
D	Roof R-20	7,500	5,700	0.760	4	−1800	0.760
E	Double glass	6,300	4,100	0.651	6	−2200	0.651
F	Reflective glass	7,520	10,000	1.330	3	2480	1.330

Note: (4) = (3)/(2), (6) = (3) − (2) and (7) = 1 + (6)/(2).

decreases benefits or savings) will affect the relative rankings of competing independent projects, and thereby influence investment decisions.[1]

- Biasing effects, detrimental to economic efficiency, can result from certain formulations of the BCR and SIR ratios. As an example, when comparing competing projects that differ significantly in their maintenance costs, placing maintenance costs in the denominator with investment costs tends to bias the final selection away from projects with relatively high maintenance costs, even though they may offer higher net benefits (profits) than competing projects. Similar biasing effects can occur in the placement of other non-investment costs such as energy or labor costs. This highlights the fact that adding a given amount to the denominator of a ratio reduces the quotient more than does subtracting an identical amount from the numerator. Hence, to eliminate this bias when the objective is to maximize the return on the investment budget, all non-investment costs should be placed in the numerator.[1]

Table 3 Projects ranked by initial cost

Project number	Initial cost	NPV	PI
1	430,000	250,500	1.583
2	380,000	215,000	1.566
3	310,000	40,000	1.129
4	260,000	60,000	1.231
5	230,000	310,000	2.348
6	185,000	105,000	1.568
7	135,000	125,000	1.926
8	110,000	76,000	1.691
9	90,000	122,000	2.356
10	85,000	96,000	2.129

PROJECT SELECTION UNDER CAPITAL RATIONING

When a firm needs to decide whether to invest in certain projects, the net present value (NPV) method is commonly used. In this method, all of the present values of both cash inflows and outflows are accounted for, and the difference in the values represents the NPV. A higher NPV reflects better rewards and higher benefits. If a firm wants to select the best of new projects within a limited budget, then it should choose the set of projects that provides the largest total NPV. Here the BCA can be a useful tool for identifying the best projects to choose under the capital rationing method.[3] This is because the BCA normalizes the evaluation as it measures the NPV per dollar invested regardless of the project size. The following two examples illustrate the selection procedure for projects under capital rationing.

Example 2

In Table 2, several energy conservation measures are assessed in terms of their economic viability. It is required to select the optimum alternative, but also to maintain the budget constraint that states that only $14,000 is available as a capital investment. Assume that the lifespan of the energy conservation measures is 15 years and the discount rate is 5%.

The BCR values are calculated utilizing the present value of money and listed in Table 2. The highest BCR value was that of the R-15 wall insulation (Alternative B), followed by R-10 wall insulation (Alternative A), and then reflective glass (Alternative F). All other measures had BCR values below 1, which means that the returns on investment in these energy conservation measures will not pay for themselves after 15 years. Due to the budget constraint, Alternatives A and F with a total cost of $13,570 represent the optimum alternatives. If the available budget was $15,000 instead of $14,000 then

Table 4 Profitability index for project selection

Project number	Initial cost	NPV	PI	Cumulative Cost	Cumulative NPV	Cumulative Cost	Cumulative NPV
9	90,000	122,000	2.36	90,000	122,000	90,000	122,000
5	230,000	310,000	2.35	320,000	432,000	320,000	432,000
10	85,000	96,000	2.13	405,000	528,000	405,000	528,000
7	135,000	125,000	1.93	540,000	653,000	540,000	653,000
8	110,000	76,000	1.69	650,000	729,000	650,000	729,000
1	430,000	250,500	1.58	1,080,000	979,500		
6	185,000	105,000	1.57	1,265,000	1,084,500	835,000	834,000
2	380,000	215,000	1.57	1,645,000	1,299,500	1,215,000	1,049,000
4	260,000	60,000	1.23	1,905,000	1,359,500	1,475,000	1,109,000
3	310,000	40,000	1.13	2,215,000	1,399,500	1,785,000	1,149,000

Alternatives B and F, with a total cost of $14,420, can be selected.

Example 3

A company has a maximum of $1.5 million to invest in retrofitting old buildings to reduce their energy consumption. After evaluating numerous projects, only those that are listed in Table 3 were selected to analyze. Unfortunately, to undertake all these projects a total fund of $2.215 million is required. Accordingly, a decision has to be made to stay within the budget constraint.

The projects in Table 4 are ranked in descending order by their PI. A higher PI is more desirable than the lower one as it creates more NPV per dollar of the initial investment.

Given this scenario (and keeping in mind the budget constraint of $1.5 million), projects 9, 5, 10, 7, 8, 1, and 6 will be selected. The total accumulative cost of these projects amounts to $1.265 million with a cumulative NPV of $1.0845 million. On the other hand, when skipping project 1 and going down the list to include projects 2 and 4, the cumulative initial cost amounts to $1.475 with an NPV of $1.109. The NPV of the second group of projects was higher than that of the first group. This means that the second group will provide a more economic solution.

CONCLUSION

Several points should be considered when conducting BCA.

- The outcome of any analysis will vary depending on the data estimates and assumptions. Thus, it is important to carefully select the assumed values for critical parameters such as the discount rate values to arrive at a realistic solution.
- If the outcome appears particularly sensitive to the value assigned to a given parameter, and the estimate is of poor or unknown quality, the analyst may wish to improve the quality of the data. Sensitivity analysis, a useful technique for identifying critical parameters, should be carried out accordingly.
- Alternatives must be compared over the same study period; otherwise, asset values including costs and savings (benefits) may be repeated.
- NPV and PI will always yield the same decision, though they will not necessarily rank projects in the same order.

REFERENCES

1. American Society for Testing and Materials (ASTM), E964. *Standard Practice for Measuring Benefit-to-Cost and Savings-to-Investment Ratios for Buildings and Building Systems*; American Society for Testing and Materials: West Conshohocken, PA, 2002.
2. American Society for Testing and Materials (ASTM), E1185. *Standard Guide for Selecting Economic Methods for Evaluating Investments in Buildings and Building Systems*; American Society for Testing and Materials: West Conshohocken, PA, 2002.
3. Emery, D.R.; Finnerty, J.D. *Corporate Financial Management*; Prentice-Hall, Inc.: Upper Saddle River, NJ, 1997.
4. Krarti, M. *Energy Audit of Building Systems, an Engineering Approach*; CRC Press: Boca Raton, FL, 2000.
5. McKenzie, A. *Benefit-Cost Analysis for Transportation Projects. Office of Investment Management*; Minnesota Department of Transportation: Washington, DC, 2005.
6. NetMBA.com. Internet Center for Management and Business Administration, Inc. *Finance: Annuities*, http://www.netmba.com/finance/time-value/annuity/ (accessed March 2006).
7. Portney, P.R. *The Concise Encyclopedia of Economics*. Benefit-Cost Analysis. http://www.econlib.org/library/Enc/BenefitCostAnalysis.html (accessed March 2006).
8. Thumann, A.; Wainwright, F. *Financing Energy Projects Deskbook*; The Fairmont Press: Atlanta, GA, 1997.

Biomass

Alberto Traverso
Dipartimento di Macchine Sistemi Energetici e Trasporti, Thermochemical Power Group, Università di Genova, Genova, Italy

Abstract
This entry deals with biomass as an energy source. Different types of biomass are described from the energy perspective, focusing on those more interesting for energy application. The main energy conversion technologies available are outlined, as well as the properties of their main products. Finally, an overview over the benefits that come from biomass exploitation for energy purposes is provided.

INTRODUCTION

This work is organized into seven main sections. The first paragraph provides the reader with a general overview on biomass, including the definition, environmental benefits, energetic properties, and a short list of biomass types that can be used as energy sources. The second paragraph illustrates the mechanical processes needed to produce standardized solid biomass fuels. The third paragraph describes one of the major technologies for converting biomass into energy, which is combustion. The fourth paragraph analyzes pyrolysis and gasification as promising techniques for efficient exploitation of biomass still in a demonstration phase. The fifth paragraph is concerned with biochemical processes for producing biogas and biofuels for transportation. The sixth paragraph outlines the major benefits from biomass exploitation for energy purposes. The seventh paragraph is constituted by the concluding remarks.

GENERALITIES ABOUT BIOMASS

In general, biomass is the substance produced or by-produced by biological processes.

Commonly, biomass refers to the organic matter derived from plants and generated through photosynthesis. Biomass not only provides food but also construction materials, fibers, medicines, and energy. In particular, biomass can be regarded as solar energy stored in the chemical bonds of the organic material. Carbon dioxide (CO_2) from the atmosphere and water absorbed by the plants roots are combined in the photosynthetic process to produce carbohydrates (or sugars) that form the biomass. The solar energy that drives photosynthesis is stored in the chemical bonds of the biomass structural components. During biomass combustion, oxygen from the atmosphere combines with the carbon and hydrogen in biomass to produce CO_2 and water. The process is therefore cyclic because the carbon dioxide is then available to produce new biomass. This is also the reason why bioenergy is potentially considered carbon-neutral, although some CO_2 emissions occur due to the use of fossil fuels during the production and transport of biofuels.

Biomass resources can be classified according to the supply sector, as shown in Table 1.

The chemical composition of plant biomass varies among species. Yet, in general terms, plants are made of approximately 25% lignin and 75% carbohydrates or sugars. The carbohydrate fraction consists of many sugar molecules linked together in long chains or polymers. Two categories are distinguished: cellulose and hemi-cellulose. The lignin fraction consists of non-sugar-type molecules that act as a glue, holding together the cellulose fibers.

The Energy Content of Biomass

Bioenergy is energy of biological and renewable origin, normally derived from purpose-grown energy crops or by-products of agriculture. Examples of bioenergy resources are wood, straw, bagasse, and organic waste. The term bioenergy encompasses the overall technical means through which biomass is produced, converted, and used. Fig. 1 summarizes the variety of processes for energy production from biomass.

The calorific value of a fuel is usually expressed as the higher heating value (HHV) or the lower heating value (LHV). The difference is caused by the heat of evaporation of the water formed from the combustion of hydrogen in the material and the original moisture. Note that the difference between the two heating values depends on the chemical composition of the fuel.

Keywords: Biomass; Renewable energy; Bioenergy; Biomass conversion; Biofuel.

Table 1 Types of biomass for energy use

Supply sector	Type	Example
Forestry	Dedicated forestry	Short rotation plantations (e.g., willow, poplar, eucalyptus)
	Forestry by-products	Wood blocks, wood chips from thinnings
Agriculture	Dry lignocellulosic energy crops	Herbaceous crops (e.g., miscanthus, reed canary-grass, giant reed)
	Oil, sugar, and starch energy crops	Oil seeds for methylesters (e.g., rape seed, sunflower)
		Sugar crops for ethanol (e.g., sugar cane, sweet sorghum)
		Starch crops for ethanol (e.g., maize, wheat)
	Agricultural residues	Straw, prunings from vineyards, and fruit trees
	Livestock waste	Wet and dry manure
Industry	Industrial residues	Industrial waste wood, sawdust from saw mills
		Fibrous vegetable waste from paper industries
Waste	Dry lignocellulosic	Residues from parks and gardens (e.g., prunings, grass)
	Contaminated waste	Demolition wood
		Organic fraction of municipal solid waste
		Biodegradable landfilled waste, landfill gas
		Sewage sludge

Source: From European Biomass Industry Association and DOE Biomass Research and Development Initiative (see Refs. 1, 2).

The most important property of biomass feedstocks with regard to combustion—and to the other thermochemical processes—is the moisture content, which influences the energy content of the fuel. Wood, just after falling, has a typical 55% water content and LHV of approximately 7.1 MJ/kg; logwood after 2–3 years of air-drying may present 20% water content and LHV of 14.4 MJ/kg; pellets show a quite constant humidity content of about 8% with LHV equal to 17 MJ/kg.

MECHANICAL PROCESSES FOR ENERGY DENSIFICATION

Some practical problems are associated with the use of biomass material (sawdust, wood chips, or agricultural residues) as fuel. Those problems are mainly related to the high bulk volume, which results in high transportation costs and requires large storage capacities, and to the high moisture content, which can result in biological degradation as well as in freezing and blocking the in-plant transportation systems. In addition, variations in moisture content make difficult the optimal plant operation and process control. All of those problems may be overcome by standardization and densification. The former consists of processing the original biomass in order to obtain fuels with standard size and heating properties, while the latter consists in compressing the material, which needs to be available in the sawdust size, to give it more uniform properties. Table 2 reports the main features of pellets, briquettes, and chips.

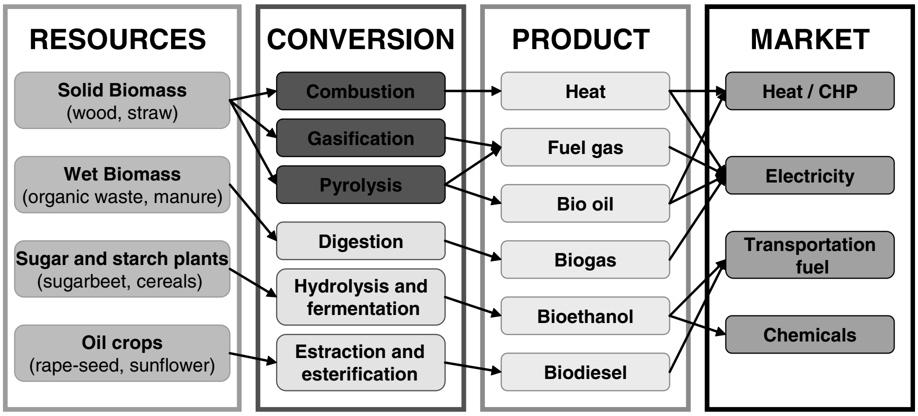


Fig. 1 Processes to convert biomass into useful energy, i.e., bioenergy. Source: From European Biomass Industry Association, Elsevier Applied Science Publishers Ltd, and Risoe National Laboratory (see Refs.1, 3, and 4).

Table 2 Comparison of different solid wood fuels

	Pellets	Briquettes	Chips
Appearance			
Raw material	Dry and ground wood or agricultural residues	Dry and ground wood or agricultural residues. Raw material can be more coarse than for pelleting due to the larger dimensions of final product	Dry wood logs
Shape	Cylindrical (generally Ø 6–12 mm, with a length 4–5 times the Ø)	Cylindrical (generally Ø 80–90 mm) or parallelepiped (150×70×60 mm)	Irregularly parallelepiped (70×30×3 mm)

BIOMASS COMBUSTION

The burning of wood and other solid biomass is the oldest energy technology used by man. Combustion is a well-established commercial technology with applications in most industrialized and developing countries, and development is concentrated on resolving environmental problems, improving the overall performance with multi-fuel operation, and increasing the efficiency of the power and heat cycles (CHP).

The devices used for direct combustion of solid biomass fuels range from small domestic stoves (1–10 kW) to the large boilers used in power and CHP plants (>5 MW). Intermediate devices cover small boilers (10–50 kW) used for heating in single-family houses, medium-sized boilers (50–150 kW) used for heating in multifamily houses or buildings, and large boilers (150 kW to over 1 MW) used for district heating. Cofiring in fossil-fired power stations enables the advantages of large size plants (>100 MWe) that are not applicable for dedicated biomass combustion due to limited local biomass availability.

To achieve complete burnout and high efficiencies in small-scale combustion, downdraft boilers with inverse flow have been introduced, which apply the two-stage combustion principle. An operation at very low load should be avoided as it can lead to high emissions. Hence, it is recommended to couple log wood boilers to a heat storage tank. Because wood pellets are well suited for automatic heating at small heat outputs (as needed for nowadays buildings), pellet furnaces are an interesting application with growing propagation. Thanks to the well-defined fuel at low water content, pellet furnaces can easily achieve high combustion quality. They are applied both as stoves and as boilers, and they are encountering increasing acceptance especially in urban areas because modern pellet stoves are nowadays efficient home heating appliances. While a conventional fireplace is less than 10% efficient at delivering heat to a house, an average modern pellet stove achieves 80%–90% efficiency. Technology development has led to the application of strongly improved heating systems that are automated and have catalytic gas cleaning equipment. Such systems significantly reduce the emissions from fireplaces and older systems while at the same time improving the efficiency.

Understoker furnaces are mostly used for wood chips and similar fuel with relatively low ash content, while grate furnaces can also be applied for high ash and water content. Special types of furnaces have been developed for straw that has very low density and is usually stored in bales. Other than conventional grate furnaces operated with whole bales, cigar burners and other specific furnaces are in operation. Stationary or bubbling fluidized bed (SFB) as well as circulating fluidized bed (CFB) boilers are applied for large-scale applications and are often used for waste wood or mixtures of wood and industrial wastes, e.g., from the pulp and paper industry.

Co-Combustion

Bioenergy production may be hampered by limitations in the supply or by fuel quality. In those cases, the cofiring of several types of biomass or cofiring biomass with coal ensures flexibility in operation, both technically and economically. Several concepts have been developed:

- *Co-combustion or direct cofiring*. The biomass is directly fed to the boiler furnace, if needed, after physical preprocessing of the biomass such as drying, grinding, or metal removal is applied. This typically takes place in bubbling or CFB combustors. Such technologies can be applied to a wide range of fuels, even for very wet fuels like bark or sludge. Multifuel

fluidized bed boilers achieve efficiencies over 90% while flue gas emissions are lower than for conventional grate combustion due to lower combustion temperatures.
- *Indirect cofiring*. Biomass is first gasified and the fuel gas is cofired in the main boiler. Sometimes the gas has to be cooled and cleaned, which is more challenging and implies higher operation costs.
- *Parallel combustion*. The biomass is burnt in a separate boiler for steam generation. The steam is used in a power plant together with the main fuel.

Problems in Biomass Combustion

Biomass has a number of characteristics that makes it more difficult to handle and combust than fossil fuels. The low energy density is the main problem in handling and transport of the biomass, while the difficulties in using biomass as fuel relates to its content of inorganic constituents. Some types of biomass used contain significant amounts of chlorine, sulfur, and potassium. The salts—KCl and K_2SO_4—are quite volatile, and the release of these components may lead to heavy deposition on heat transfer surfaces, resulting in reduced heat transfer and enhanced corrosion rates. Severe deposits may interfere with operation and cause unscheduled shut downs.

In order to minimize these problems, various fuel pretreatment processes have been considered, including washing the biomass with hot water or using a combination of pyrolysis and char treatment.

THERMOCHEMICAL CONVERSION OF BIOMASS

Pyrolysis and gasification are the two most typical thermochemical processes that do not produce useful energy directly because they convert the original bioenergy feedstock into more convenient energy carriers such as producer gas, oil, methanol, and char.[3]

Pyrolysis

Pyrolysis is a process for thermal conversion of solid fuels, like biomass or wastes, in the complete absence of oxidizing agent (air/oxygen) or with such limited supply that gasification does not occur to any appreciable extent. Commercial applications are either focused on the production of charcoal or the production of a liquid product—the bio-oil and pyro-gas. Charcoal is a very ancient product, even if traditional processes (partial combustion of wood covered by a layer of earth) are very inefficient and polluting. Modern processes such as rotary kiln carbonization are presently used in industry. Bio-oil production (or wood liquefaction) is potentially very interesting as a substitute for fuel oil and as a feedstock for production of synthetic gasoline or diesel fuel. Pyro-gas has higher energy density than gasification gas (syngas) because it has been created without oxygen (and nitrogen, if air is employed), hence it does not contain the gaseous products of partial combustion.

The pyrolysis process takes place at temperatures in the range 400°C–800°C and during this process most of the cellulose and hemicellulose and part of the lignin will disintegrate to form smaller and lighter molecules, which are gases at the pyrolysis temperature. As these gases cool, some of the vapors condense to form a liquid, which is the bio-oil and the tars. The remaining part of the biomass, mainly parts of the lignin, is left as a solid, i.e., the charcoal. It is possible to influence the product mix through a control of heating rate, residence time, pressure, and maximum reaction temperature so that either gases, condensable vapors, or the solid charcoal is promoted.

Gasification

Gasification technology has been developing since the 18th century, and it is still in a development phase.[5,6] Gasification is a conversion process that involves partial oxidation at elevated temperature. It is intermediate between combustion and pyrolysis. In fact, oxygen (or air) is present but it is not enough for complete combustion. This process can start from carbonaceous feedstock such as biomass or coal and convert them into a gaseous energy carrier. The overall gasification process may be split into two main stages: the first is pyrolysis stage, i.e., where oxygen is not present but temperature is high, and here typical pyrolysis reactions take place; the second stage is the partial combustion, where oxygen is present and it reacts with the pyrolyzed biomass to release heat necessary for the process. In the latter stage, the actual gasification reactions take place, which consist of almost complete charcoal conversion into lighter gaseous products (e.g., carbon monoxide and hydrogen) through the chemical oxidizing action of oxygen, steam, and carbon dioxide. Such gases are injected into the reactor near the partial combustion zone (normally, steam and carbon dioxide are mutually exclusive). Gasification reactions require temperature in excess of 800°C to minimize tar and maximize gas production. The gasification output gas, called "producer gas," is composed by hydrogen (18%–20%), carbon monoxide (18%–20%), carbon dioxide (8%–10%), methane (2%–3%), trace amounts of higher hydrocarbons like ethane and ethene, water, nitrogen (if air is used as oxidant agent), and various contaminants such as small char particles, ash, tars, and oils. The incondensable part of producer gas is called "syngas" and it represents the useful product of gasification. If air is used, syngas has a high heating

Bene–Clim

value in the order of 4–7 MJ/m^3, which is exploitable for boiler, engine, and turbine operation, but due to its low energy density, it is not suitable for pipeline transportation. If pure oxygen is used, the syngas high heating value almost doubles (approximately 10–18 MJ/m^3 high heating value), hence such a syngas is suitable for limited pipeline distribution as well as for conversion to liquid fuels (e.g., methanol and gasoline). However, the most common technology is the air gasification because it avoids the costs and the hazards of oxygen production and usage. With air gasification, the syngas efficiency—describing the energy content of the cold gas stream in relation to that of the input biomass stream—is in the order of 55%–85%, typically 70%.

Comparison of Thermal Conversion Methods of Biomass

Table 3 reports a general overview on specific features of the conversion technologies analyzed here, showing the related advantages and drawbacks.

BIOCHEMICAL CONVERSION OF BIOMASS

Biochemical conversion of biomass refers to processes that decompose the original biomass into useful products. Commonly, the energy product is either in the liquid or in gaseous forms, hence it is called "biofuel" or "biogas," respectively. Biofuels are very promising for transportation sector, while biogas is used for electricity and heat production. Normally, biofuels are obtained from dedicated crops (e.g., biodiesel from seed oil), while biogas production results from concerns over environmental issues such as the elimination of pollution, the treatment of waste, and the control of landfill greenhouse gas emissions.

Biogas from Anaerobic Digestion

The most common way to produce biogas is anaerobic digestion of biomass. Anaerobic digestion is the bacterial breakdown of organic materials in the absence of oxygen. This biochemical process produces a gas called biogas, principally composed of methane (30%–60% in volume) and carbon dioxide. Such a biogas can be converted to energy in the following ways:

- Biogas converted by conventional boilers for heating purposes at the production plant (house heating, district heating, industrial purposes).
- Biogas for combined heat and power generation.
- Biogas and natural gas combinations and integration in the natural gas grid.
- Biogas upgraded and used as vehicle fuel in the transportation sector.
- Biogas utilization for hydrogen production and fuel cells.

An important production of biogas comes from landfills. Anaerobic digestion in landfills is brought about by the microbial decomposition of the organic matter in refuse. Landfill gas is on average 55% methane and 45% carbon dioxide. With waste generation increasing at a faster rate than economic growth, it makes sense to recover the energy from that stream through thermal or fermentation processes.

Biofuels for Transport

A wide range of chemical processes may be employed to produce liquid fuels from biomass. Such fuels can find a very high level of acceptance by the market thanks to their relatively easy adaptation to existing technologies (i.e., gasoline and diesel engines). The main potential biofuels are outlined below.

- Biodiesel is a methyl-ester produced from vegetable or animal oil to be used as an alternative to conventional petroleum-derived diesel fuel. Compared to pure vegetable or animal oil, which can be used in adapted diesel engines as well, biodiesel presents lower viscosity and slightly higher HHV.
- Pure vegetable oil is produced from oil plants through pressing, extraction, or comparable procedures, crude or refined but chemically unmodified. Usually, it is

Table 3 Qualitative comparison of technologies for energy conversion of biomass

Process	Technology	Economics	Environment	Market potential	Present deployment
Combustion–heat	+++	€	+++	+++	+++
Combustion–electricity	++(+)	€€	++(+)	+++	++
Gasification	+(+)	€€€	+(++)	+++	(+)
Pyrolysis	(+)	€€€€	(+++)	++(+)	(+)

+, low; +++, high; €, cheap; €€€€, expensive.
Source: From European Biomass Industry Association and Risoe National Laboratory (see Refs. 1, 4).

compatible with existing diesel engines only if blended with conventional diesel fuel at rates not higher than 5%–10% in volume. Higher rates may lead to emission and engine durability problems.

- Bioethanol is ethanol produced from biomass or the biodegradable fraction of waste. Bioethanol can be produced from any biological feedstock that contains appreciable amounts of sugar or other matter that can be converted into sugar, such as starch or cellulose. Also, ligno-cellulosic materials (wood and straw) can be used, but their processing into bioethanol is more expensive. Application is possible to modified spark ignition engines.
- Bio-ETBE (ethyl-tertio-butyl-ether) is ETBE produced on the basis of bioethanol. Bio-ETBE may be effectively used to enhance the octane number of gasoline (blends with petrol gasoline).
- Biomethanol is methanol produced from biomass. Methanol can be produced from gasification syngas (a mixture of carbon monoxide and hydrogen) or wood dry distillation (old method with low methanol yields). Most all syngas for conventional methanol production is produced by the steam reforming of natural gas into syngas. In the case of biomethanol, a biomass is gasified first to produce a syngas from which the biomethanol is produced. Application is possible to spark ignition engines and fuel cells. Compared to ethanol, methanol presents more serious handling issues because it is corrosive and poisonous for human beings.
- Bio-MTBE (methyl-tertio-butyl-ether) is a fuel produced on the basis of biomethanol. It is suitable for blends with petrol gasoline.
- Biodimethylether (DME) is dimethylether produced from biomass. Bio-DME can be formed from syngas by means of oxygenate synthesis. It has emerged only recently as an automotive fuel option. Storage capabilities are similar to those of LPG. Application is possible to spark ignition engines.

BENEFITS FROM BIOMASS ENERGY

There is quite a wide consensus that, over the coming decades, modern biofuels will provide a substantial source of alternative energy. Nowadays, biomass already provides approximately 11%–14% of the world's primary energy consumption (data varies according to sources).

There are significant differences between industrialized and developing countries. In particular, in many developing countries, bioenergy is the main energy source—even if used in very low-efficiency applications (e.g., cooking stoves have an efficiency of about 5%–15%). Furthermore, inefficient biomass utilization is often associated with the increasing scarcity of hand-gathered wood, nutrient depletion, and the problems of deforestation and desertification.

One of the key drivers to bioenergy deployment is its positive environmental benefit, in particular regarding the global balance of green house gas (GHG) emissions. This is not a trivial matter, because biomass production and use are not entirely GHG neutral. In general terms, the GHG emission reduction as a result of employing biomass for energy reads as reported in Table 4.

Bioenergy is a decentralized energy option whose implementation presents positive impacts on rural development by creating business and employment opportunities. Jobs are created all along the bioenergy chain, from biomass production or procurement to its transport, conversion, distribution, and marketing.

Bioenergy is a key factor for the transition to a more sustainable development.

CONCLUSIONS

Biomass refers to a very wide range of substances produced by biological processes. In the energy field, special focus has been and will be placed on vegetable biomass such as wood and agricultural by-products because of the energy potential as well as economic and environmental benefits. Size and humidity standardization of biomass is a necessary step to make it suitable for effective domestic and industrial exploitation. Chips, briquettes, and pellets are modern examples of standard solid fuels.

Biomass can be converted to energy in three may pathways: combustion, thermochemical processing, and biochemical processing. The combustion of solid biomass for the production of heat or electricity is the most viable technology, while pyrolysis and gasification still face economic and reliability issues. Among biochemical processes, anaerobic digestion is often used to reduce the environmental impact of hazardous waste and landfills. Biochemical processes are also

Table 4 Benefits in reduction of green houses gas emissions

+	Avoided mining of fossil resources
−	Emission from biomass production
+	Avoided fossil fuel transport (from producer to user)
−	Emission from biomass fuel transport (from producer to user)
+	Avoided fossil fuel utilization

+, positive; −, neutral.
Source: From Risoe National Laboratory (see Ref. 4).

Bene–Clim

concerned with the conversion of biomass into useful fuels for transportation, such as biodiesel, bioethanol, biomethanol, and others. All of them can effectively contribute to the transition to a more sustainable transportation system at zero GHG emissions.

Biomass represents a viable option for green energy resources of the 21st century.

ACKNOWLEDGMENTS

The author wishes to thank Professor A. F. Massardo of the University of Genoa for the invaluable help during the last years of research activity.

REFERENCES

1. European Biomass Industry Association, http://www.eubia.org (accessed on November 21, 2006).
2. DOE Biomass Research and Development Initiative, http://www.brdisolutions.com (accessed on November 21, 2006).
3. Overend, R.P.; Milne, T.A.; Mudge, L.K. *Fundamentals of Thermochemical Biomass Conversion*; Elsevier Applied Science Publishers Ltd: NewYork, 1985.
4. Risoe National Laboratory, Denmark, http://www.risoe.dk (accessed on November 21, 2006).
5. Bridgewater, A.V. The technical and economic feasibility of biomass gasification for power generation. Fuel J. **1995**, *74* (6–8), 557–564.
6. Franco, A.; Giannini, N. Perspectives for the use of biomass as fuel in combined cycle power plants. Int. J. Thermal Sci. **2005**, *44* (2), 163–177.

Boilers and Boiler Control Systems

Eric Peterschmidt
Honeywell Building Solutions, Honeywell, Golden Valley, Minnesota, U.S.A.
Michael Taylor
Honeywell Building Solutions, Honeywell, St. Louis Park, Minnesota, U.S.A.

Bene–Clim

Abstract
Many commercial and industrial facilities use boilers to produce steam or hot water for space heating or for process heating. Boilers are typically major users of energy, and any person involved in energy management needs to know how a boiler works and how the performance of a boiler can be maintained or improved. This article outlines the types of boilers used to heat facilities, while providing an overview of basic boiler controls and the parameters that affect energy efficiency.

INTRODUCTION

A boiler is a closed vessel intended to heat water and produce hot water or steam through the combustion of a fuel or through the action of electrodes or electric resistance elements. Many commercial and industrial facilities use boilers to produce steam or hot water for space heating or for process heating. Boilers are typically major users of energy, and any person involved in a facility's energy management needs to know how a boiler works and how the performance of a boiler can be maintained or improved. In particular, it is important to know what parameters of a boiler system are the most important. For fossil fuel-fired boilers, the combustion efficiency is the major parameter of interest; this is most often controlled by providing the optimum amount of combustion air that is mixed with the fuel. Thus, understanding boiler control systems is extremely important. Steam and hot water boilers are available in standard sizes from very small boilers for apartments and residences to very large boilers for commercial and industrial uses.

BOILER TYPES

Boilers are classified by water temperature or steam pressure. They are further classified by the type of metal used in construction (cast iron, steel, or copper), by the type of fuel or heat element (oil, gas, or electricity), or by the relationship of fire or water to the tubes (i.e., firetube or watertube).

– Low-pressure boilers are those designed to produce steam up to 15 psig or hot water up to 250°F with pressures up to 160 psig.
– Medium- and high-pressure boilers produce steam above 15 psig or hot water above 160 psig or 250°F or both.

Boilers are typically constructed of cast iron or welded steel. Cast iron boilers (Fig. 1) are made of individually cast sections and are joined together using screws or nuts and tie rods or threaded rivets. The number of sections can be varied to provide a range of capacities.

Steel boilers come in a wide variety of configurations. They are factory-assembled and welded and shipped as a unit. Fig. 2 illustrates a firetube boiler. The fire and flue gases are substantially surrounded by water. The products of combustion pass through tubes to the back then to the front and once more to the back before finally exiting at the front. This makes it a four-pass boiler. Firetube boilers are manufactured in many other configurations such as:

– External firebox—The firebox is not surrounded by water.
– Dry back—Firetubes are directly available from clean-out doors at the back of boiler.
– Scotch–Marine—Employs low water volume and has a fast response.

Watertube boilers are steel body boilers used for high capacity requirements of more than 2 million Btu per hour (Btu/h). Watertube boilers use a water-cooled firebox which prolongs the life of furnace walls and refractories.

Keywords: Boiler; Boiler controls; Automatic temperature controls; Water/steam; Combustion; Heating; Safeguard control; Energy efficiency.

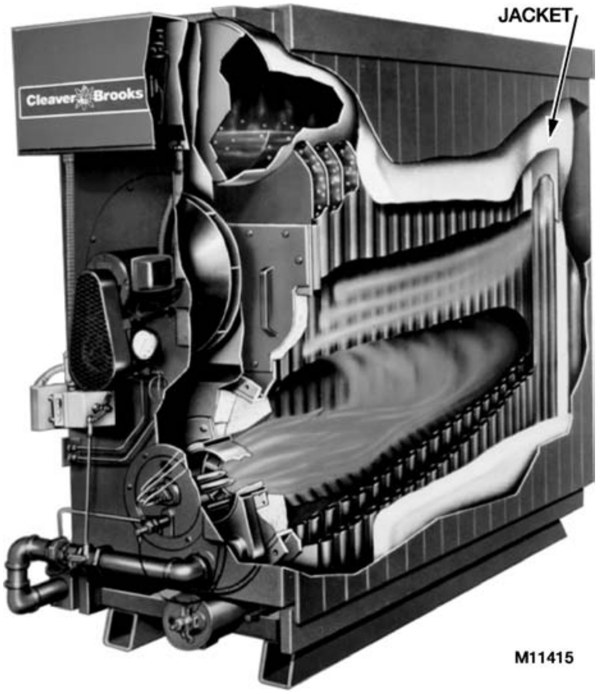


Fig. 1 Typical cast iron boiler (watertube).

Modular boilers are small, hot water boilers rated from 200,000 to 900,000 Btu/h input. These boilers are available with gross efficiencies of 85% or higher. Fig. 3 shows the features of a typical modular boiler. These boilers are often used in tandem to provide hot water for space heating and/or domestic hot water. For example, if the designed heating load were 2 million Btu/h, four 600,000 Btu/h (input) modular boilers might be used. If the load were 25% or less on a particular day, only one boiler would fire and cycle on and off to supply the load. The other three boilers would remain off with no water flow. This reduces the flue and jacket (covering of the boiler) heat losses.

Some modular boilers have a very small storage capacity and very rapid heat transfer so water flow must be proven before the burner is started.

Electric boilers heat water or produce steam by converting electrical energy to heat using either resistance elements or electrodes. Electric boilers are considered to be 100% efficient since all power that is consumed directly produces hot water or steam. Heat losses through the jacket and insulation are negligible and there is no flue.[1]

Electrode boilers (as seen in Fig. 4) have electrodes immersed in the water. Electrical current passes through the water between electrodes, and this current and the resistance of the water results in generated heat. Electrode boilers are available in sizes up to 11,000 kW. Resistance boilers have the resistance (heating) elements immersed in, but electrically insulated from, the water and are manufactured in sizes up to 3,000 kW. Electric elements and electrodes are usually grouped to provide four or more stages of heating. A step controller responds to steam pressure or hot water temperature, activating each stage of heating as required to heat the building.

BOILER RATINGS AND EFFICIENCY

Boilers can be rated in several ways. Fig. 5 shows the commonly used ratings and terms. The terms Btu/h (Btu per hour) and MBtu/h or MB/H (1000 Btu/h) indicate the

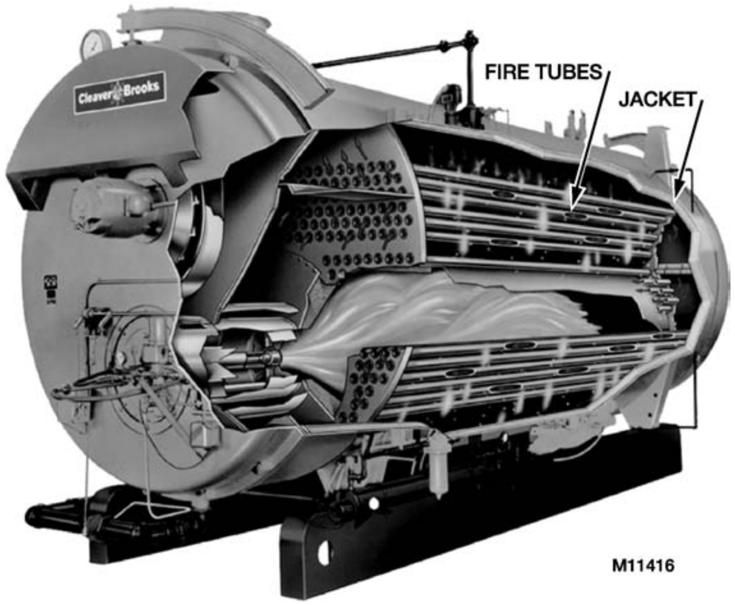


Fig. 2 Typical firetube boiler.

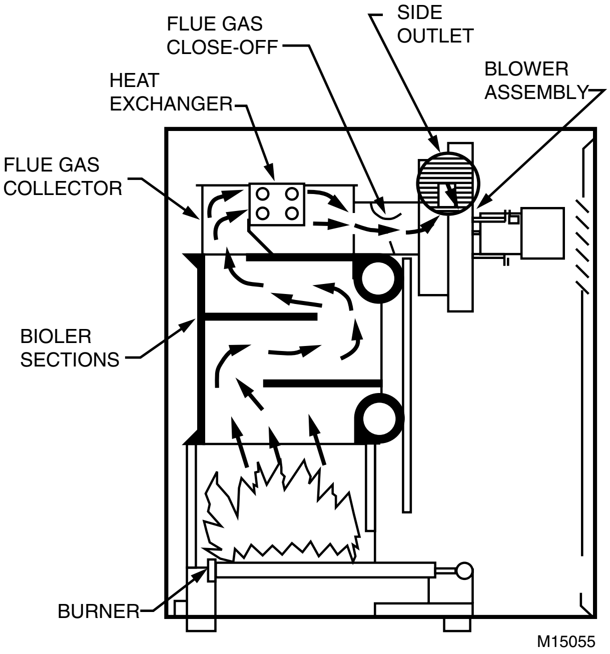


Fig. 3 High efficiency modular boiler.

boiler's input rate. Input ratings are usually shown on the boiler's (or burner's) nameplate. The terms bhp (boiler horse power), EDR (equivalent direct radiation), and pounds per hour (of steam) indicate the boiler's output rate.

Gross efficiency of the boiler is the output (steam or water heat content and volume) divided by the fuel input (measured by a fuel meter at steady-state firing conditions). The combustion efficiency, as indicated by flue gas conditions, does not take into account jacket, piping, and other losses, so it is always higher than the gross efficiency.

A testing procedure issued by the U.S. Department of Energy in 1978 measures both on-cycle and off-cycle losses based on a laboratory procedure involving cyclic conditions. The result is called the AFUE (Annual Fuel Utilization Efficiency) rating, or seasonal efficiency, which is lower than gross efficiency.

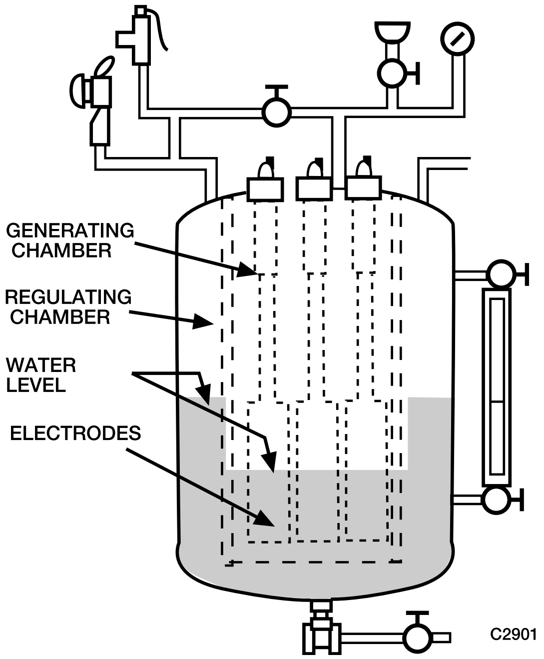


Fig. 4 Electrode steam boiler.

COMBUSTION IN BOILERS

When gas, oil, or another fuel is burned, several factors must be considered if the burning process is to be safe, efficient, and not impact the environment. The burning process must adhere to the following guidelines:

1. Provide enough air so that combustion is complete, and undesirable amounts of carbon monoxide or other pollutants are not generated.
2. Avoid excess air in the fuel-air mixture which would result in low efficiency.
3. Completely mix the air with fuel before introducing the mixture into the firebox.
4. Provide safety controls so that fuel is not introduced without the presence of an ignition flame or spark and so that flame is not introduced in the presence of unburned fuel.
5. Avoid water temperatures below the dewpoint of the flue gas to prevent condensation on the fireside of the boiler.

Combustion can be monitored by flue gas analysis. For large boilers, over 1,000,000 Btu/h, the analysis is typically continuous. For small boilers, flue gas is analyzed periodically using portable instruments. Flue gas composition analysis routinely measures the percentage of CO_2 (carbon dioxide) or O_2 (oxygen), but generally not both. Ideal CO_2 concentration is in the 10%–12% range. The percentage of oxygen remaining is the most reliable indication of complete combustion. The ideal O_2 concentration in the flue gas is in the 3%–5% range. Lower concentrations are impractical and often unsafe. Higher O_2 concentrations mean that an excessive quantity of air is being admitted to the combustion chamber and must be heated by the fuel. This excess air passes through the boiler too quickly for the heat to be efficiently transferred to the water or steam, and thereby reduces the combustion efficiency. CO_2 measuring instruments are simpler and cost less than O_2 measuring instruments.

The CO_2 or O_2 concentration, plus the stack temperature, provides a burner combustion efficiency in percent—either directly or by means of charting. This combustion efficiency indicates only the amount of heat extracted from

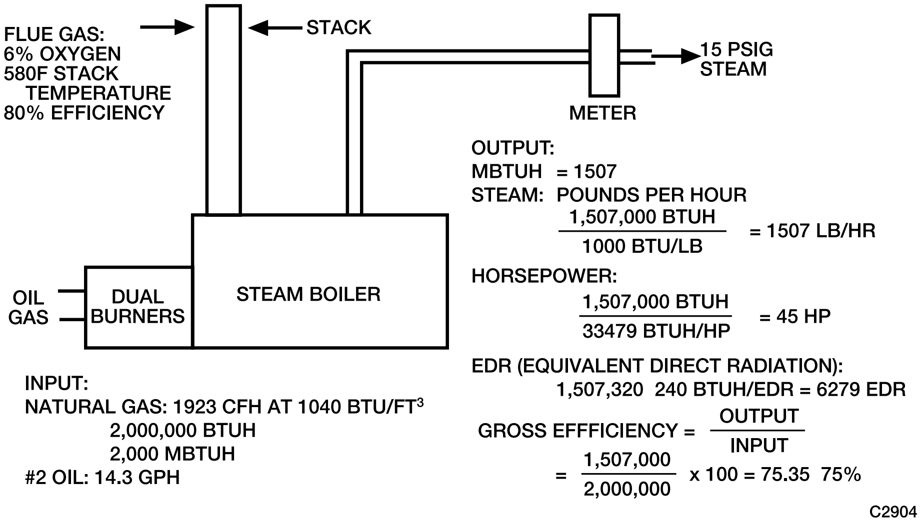


Fig. 5 Boiler ratings and efficiency.

the fuel. It does not account for excess heating of combustion air, or losses from leaks or the boiler jacket, among other factors.

For oil-fired boilers, the oil burners are usually of the atomizing variety, that is, they provide a fine spray of oil. Several types of these oil burners exist:

- Gun type burners spray oil into a swirling air supply.
- Horizontal, rotary burners use a spinning cup to whirl oil and air into the furnace.
- Steam- or air-atomizing burners use high pressured air or 25 psig steam to break up the oil into fine droplets.

For modulating or high/low flame control applications, the rotary or steam/air-atomizing burners are most common.

For natural gas-fired boilers, the two typical types of gas burners are the atmospheric injection burner and the power type burner. The atmospheric injection burner uses a jet of gas to aspirate combustion air and is commonly used in home gas furnaces and boilers. The raw-gas ring burner (refer to Fig. 6) is an atmospheric injection burner. Power burners (refer to Fig. 7) use a forced-draft fan to thoroughly mix air and gas as they enter the furnace. Common power burner applications are in the commercial and industrial sectors.

BASIC BOILER CONTROLS

Boilers have to provide steam or hot water whenever heat is needed. A conventional BMCS (boiler management control system) is often set to provide a continuous hot water or steam supply between October and May at anytime the OA (outside air) temperature drops to 60°F for more than 30 min and an AHU (air handling unit) is calling for heat. The BCMS should include a software on/off/auto function. Unlike chillers, boilers can be left enabled at no-load conditions, during which time the water temperature will be held at the designed temperature. Frequent warm-up and shut-down of boilers causes stress buildup. Boiler manufacturers' recommendations provide specific guidelines in this area of operation.

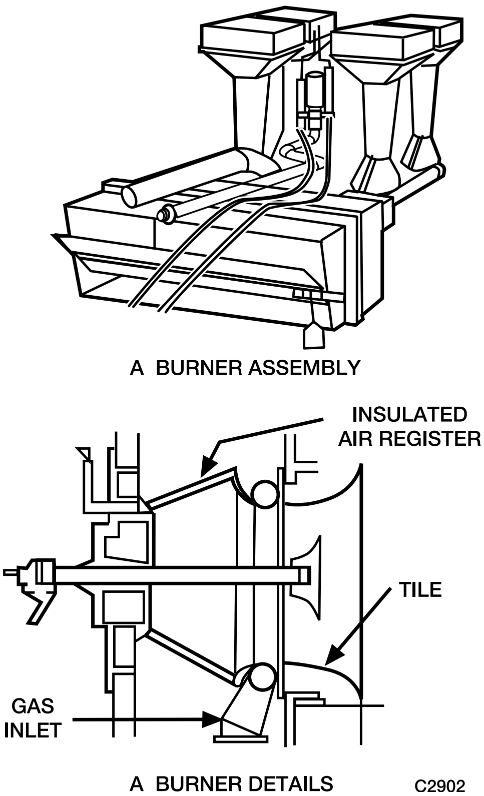


Fig. 6 Raw gas ring burner.

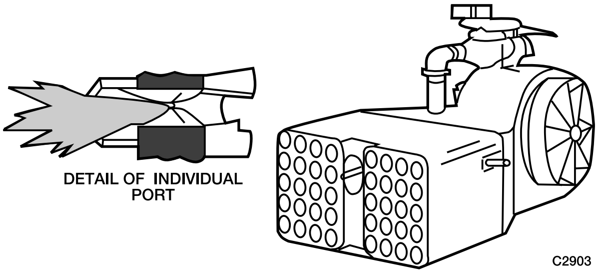


Fig. 7 Multiport forced-draft gas burner.

Unless a low-limit for water temperature is used, hot water boiler burners are not controlled to provide water temperatures based on outdoor temperatures, because the reset schedules require water temperatures to be supplied below the dewpoint temperature of the flue gas. Some boilers require incoming water temperatures to be above 140°F before going to high-fire. In this case, if a building is using a hot water system and the boiler is locked into low-fire because the incoming water is too cold, the system may never recover.

The following are three ways to control the output of a commercial boiler:

1. On/off (cycling) control
2. High-fire/low-fire control
3. Modulating control

On/off (cycling) control is most common for small boilers up to 1,000,000 Btu/h capacity. The oil or gas burner cycles on and off to maintain steam pressure or water temperature. Cycling control causes losses in efficiency because of the cooling (which is necessary for safety) of the fireside surfaces by the natural draft from the stack during the off, pre-purge and post-purge cycles.

High-fire/low-fire burners provide fewer off-cycle losses since the burner shuts off only when loads are below the low-fire rate of fuel input.

Modulating control is used on most large boilers because it adjusts the output to match the load whenever the load is greater than the low-fire limit, which is usually not less than 15% of the full load capacity. Steam pressure or hot water temperature is measured to determine the volume of gas or oil admitted to the burner.

Boiler firing and safety controls are boiler-manufacturer furnished and code approved. A BMCS usually enables a boiler to fire, provides a setpoint, controls pumps and blending valves, and monitors operation and alarms.

Combustion control regulates the air supplied to a burner to maintain a high gross efficiency in the combustion process. More sophisticated systems use an oxygen sensor in the stack to control the amount of combustion air supplied. Smoke density detection devices can be used in the stack to limit the reduction of air so stack gases stay within smoke density limits. A continuous reading and/or recording of flue gas conditions—O_2 concentration percentage, stack temperature—is usually included in the control package of large boilers.

A simple combustion control system contains a linkage that readjusts the air supply from the same modulating motor that adjusts the fuel supply (refer to Fig. 8). There may be a provision to stop the flow of air through the fluebox during the off-cycles.

Flame Safeguard Control

Flame safeguard controls are required on all burners. Flame controls for large burners can be very complicated while controls for small burners such as a residential furnace are relatively simple. The controls must provide foolproof operation—i.e., they must make it difficult or impossible to override any of the safety features of the system. The controls also should be continuously self checked. For commercial and industrial burners, the flame

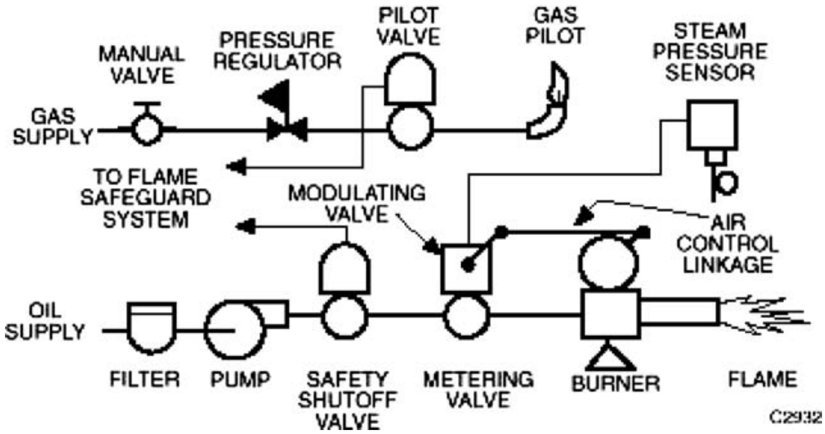


Fig. 8 Combustion control for rotary oil burner.

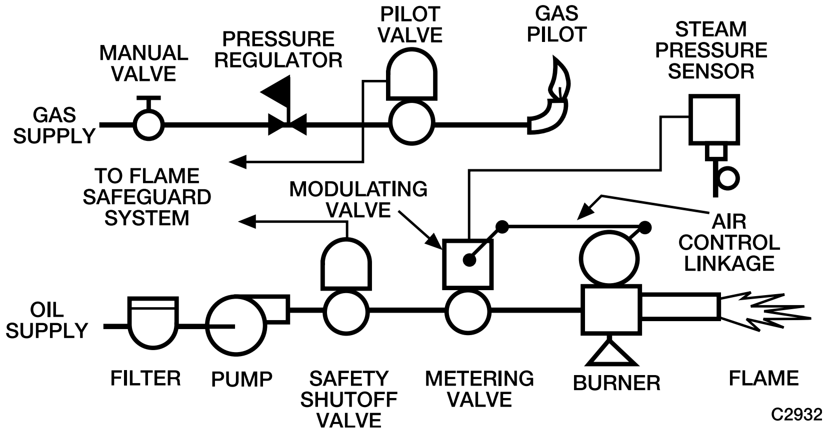


Fig. 9 Simple flame safeguard for a gas furnace.

safeguard control generally goes through a series of operations similar to the following.

- Purge the firebox of unburned fuel vapor (prepurge).
- Light the pilot.
- Verify that the pilot is lit.
- Open the main fuel valve.
- Verify that the flame is present as soon as fuel is introduced.
- Cut off the fuel supply promptly if the flame fails.
- Purge the firebox of any unburned fuel after each on-cycle (post-purge).

The key to any flame safeguard system is a reliable and fast means of detecting the presence or absence of a flame. Methods of detection include:

- Response of a bimetal sensor to heat (slow response).
- Response of a thermocouple to heat (slow response).
- Flame conductivity (fast, but not reliable response)
- Flame rectification (fast, reliable response).
- Ultraviolet flame detection (fast, reliable response).
- Lead sulfide (photo) cells (fast, reliable response if a flame frequency check is included).

Some sensors can potentially malfunction because of short circuits, hot refractories, or external light sources. Other sensors, like flame rectification and ultraviolet detection, respond to flame only. Flame safeguard systems must be approved by Underwriter's Laboratory (UL) or Factory Mutual for specific applications. Fig. 9 shows a flame safeguard system commonly applied to small gas boilers or furnaces. The flame of the gas pilot impinges on a thermocouple which supplies an electric current to keep the pilotstat gas valve open. If the pilot goes out or the thermocouple fails, the pilotstat valve closes or remains closed preventing gas flow to the main burner and pilot. The pilotstat must be manually reset.

Fig. 10 shows how flame safeguard controls are integrated with combustion controls on a small, oil-fired steam boiler. The ultraviolet (UV) flame detector is located where it can see the flame and will shutdown the burner when no flame is present.

In addition to the combustion, safety, and flame safeguard controls shown in Fig. 10, larger burners often provide additional measuring instrumentation such as:

- Percentage of O_2 or CO_2 in flue gas (to monitor combustion efficiency)
- Flue gas temperature
- Furnace draft (in inches of water) column
- Steam flow with totalizer or hot water Btu with totalizer
- Oil and/or gas flow with totalizer
- Stack smoke density

CONTROL OF MULTIPLE BOILER SYSTEMS

Basic boiler connections for a three-zone hot water system are shown in Fig. 11. In this system, two boilers are connected in parallel. Hot water from the top of the boilers moves to the air separator which removes any entrapped air from the water. The expansion tank connected to the separator maintains the pressure in the system. The tank is about half full of water under normal operating conditions. Air pressure in the tank keeps the system pressurized and allows the water to expand and contract as the system water temperature varies. Water from the boiler moves through the separator to the three zone pumps, each of which is controlled by its own thermostat. In some systems, each zone may have a central pump and a valve. Return water from each zone goes back to the boiler in the return line. Several variations are possible within this type system, but the process is the same. There is no minimum boiler water flow limit in this example.

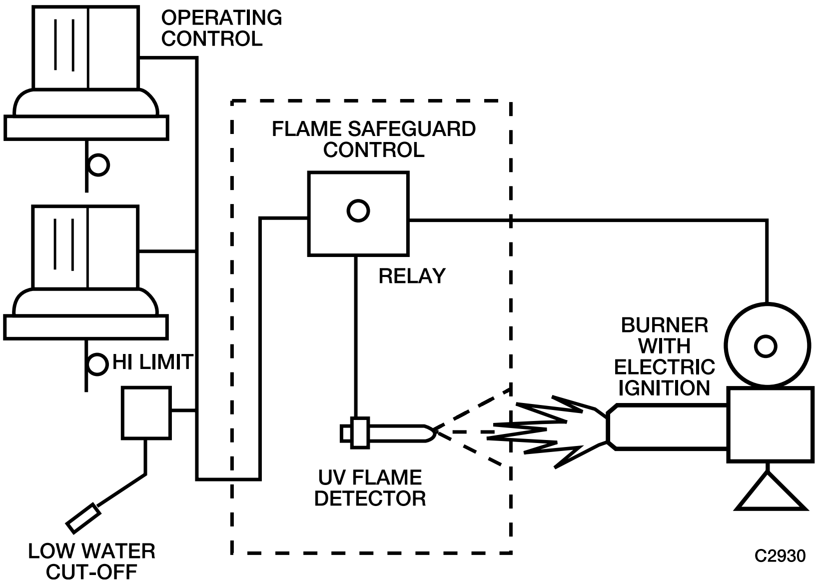
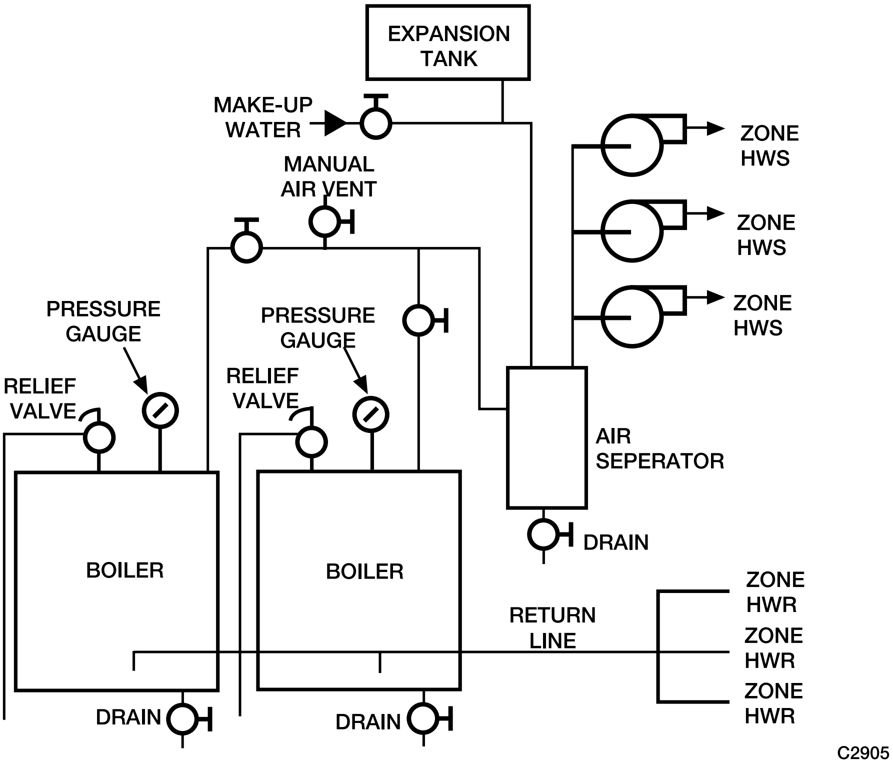


Fig. 10 Combustion controls with flame safeguard circuit.

The Dual Boiler Plant Control example in Fig. 12 is a dual boiler plant with high-fire/low-fire controlled boilers. A minimum incoming water temperature of 145°F is required prior to high-fire, water flow must be maintained when the boiler is enabled, and a secondary hot water reset schedule of 110°F water at 55°F OA temperature and 180°F water at 5°F OA temperature. These concepts adapt well for single- or multiple-boiler systems.



Note: The primary/secondary decoupler is sized for the full secondary flow, and like the chiller plant decoupler, should be a minimum of 6 pipe diameters in length. Unlike the chiller decoupler, normal flow may occur in either direction.

Fig. 11 Typical piping for multiple-zone heating system.

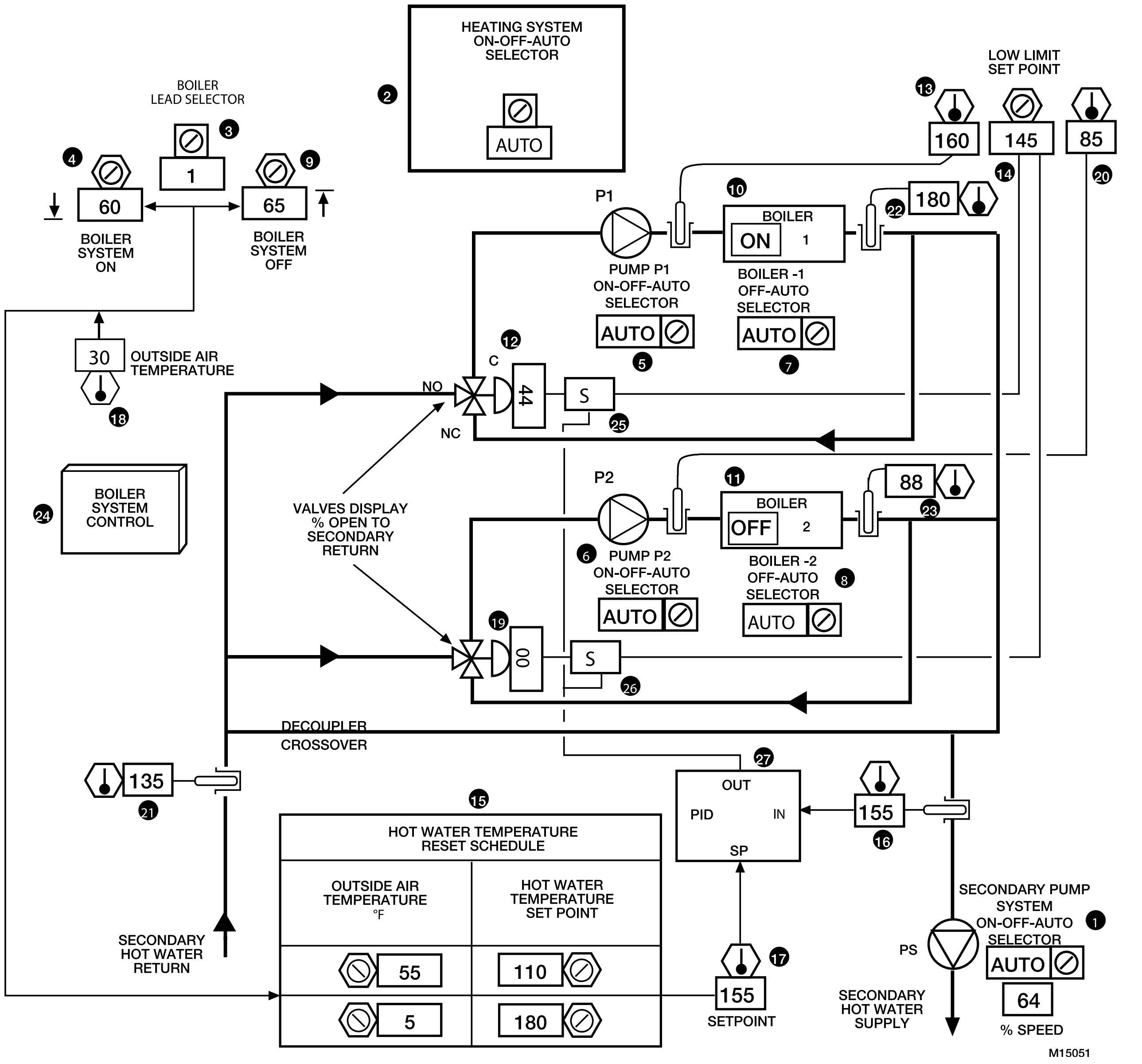


Fig. 12 Dual boiler plant control graphic.

Functional Description

Item no.	Function
1	On/off/auto function for secondary pumping system
2	On/off/auto function for heating system
3	Selects the lead boiler
4	Heating system start point (OA temperature)
5, 6	On/off/auto function for primary pumps
7, 8	Off/auto function for boilers
9	Heating system stop point (OA temperature)
10, 11	Operator information
12–14	Valve modulates to prevent incoming water from dropping below the low-limit setpoint (145°F)

Item no.	Function
15–18	Secondary water setpoint reset from OA
19, 20	Valve modulates to prevent incoming water from dropping below the low-limit setpoint (145°F)
21–23	Operator information
24	Icon, selects the Boiler System Control dynamic display (as seen in Fig. 13)
25, 26	Software signal selection functions, allows valve to control secondary HW temperature, subject to boiler low-limits
27	OA reset valve control PID

Features

1. Full flow through operating boilers
2. Minimum temperature limit on the boiler's incoming water
3. Variable-flow secondary system with full boiler flow
4. Automatic boiler staging
5. User-friendly monitoring and adjustment

Conditions for Successful Operation

1. Control network, software, and programming to advise the heating plant controller of secondary fan and water flow demands.
2. Interlock and control wiring coordinated with the boiler manufacturer.
3. Control in accord with boiler manufacturer's recommendations.
4. Proper setpoint and parameter-project-specific settings.

Specification

The heating plant shall operate under automatic control anytime the secondary pump's on/off/auto function is not "OFF," subject to a heating system's on/off/auto software function. The lead boiler, as determined by a software-driven lead-boiler-selection function, shall be enabled anytime the date is between October 1 and May 1, the OA temperature drops below 60°F for greater than 30 min, and an AHU is calling for heat. Each boiler's primary pump shall have a software on/off/auto function, and each boiler shall have a software auto/off function. The heating

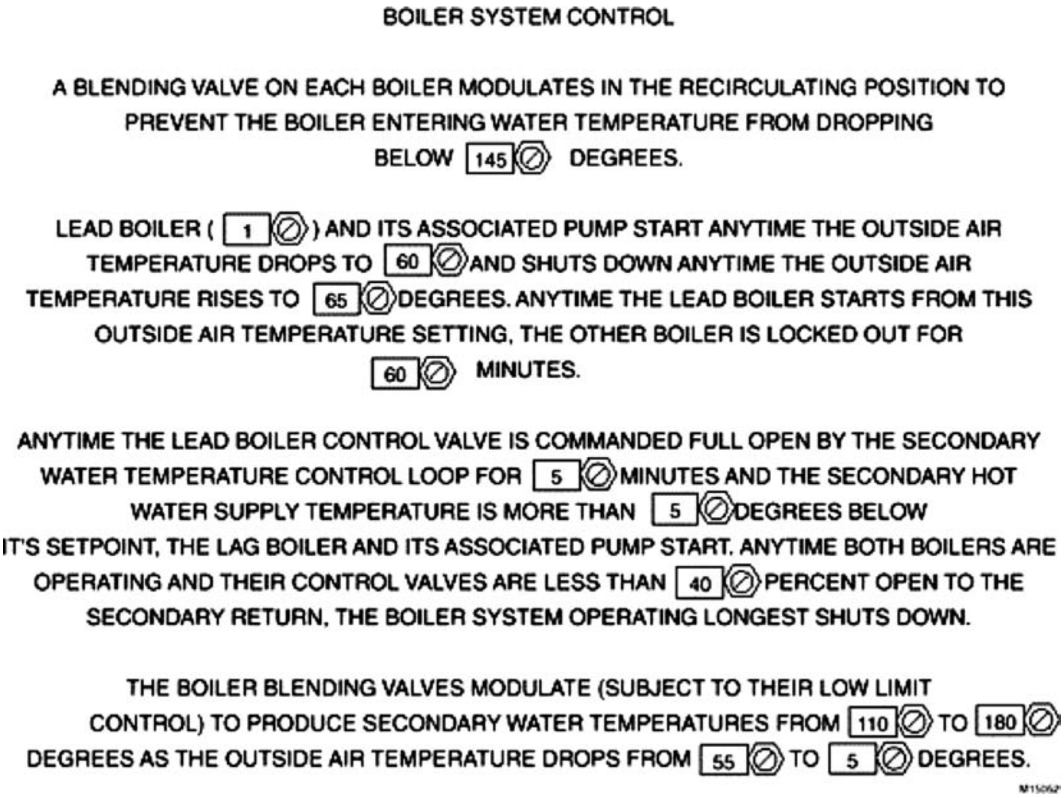


Fig. 13 Boiler system control dynamic display.

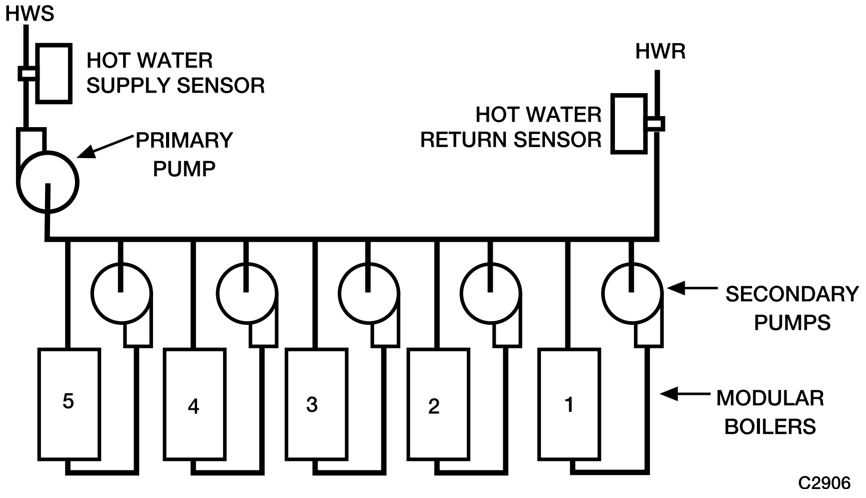


Fig. 14 Typical primary–secondary piping for modular boilers.

plant shall be disabled anytime the OA temperature rises to 65°F for greater than 1 min and after May 1.

Anytime the boiler plant is enabled, the lead boiler's primary pump shall start and, as flow is proven, the boiler shall fire under its factory controls to maintain 180°F. If the lead boiler's status does not change to "on," or if flow is not proven within 5 min, the lag boiler shall be enabled.

During boiler operation, a three-way blending valve shall position to place the boiler flow in a recirculating mode until the water entering the boiler exceeds a low-limit value of 145°F, at which time the blending valve shall modulate to maintain the secondary water temperature between 110 and 180°F as the OA temperature varies from 55 to 5°F.

The lag boiler shall be locked out from operation for 60 min after the lead boiler starts. Thereafter, anytime one boiler control valve is commanded full open by the secondary temperature control loop for greater than 5 min and the secondary water temperature is a temperature less than 5°F below the secondary water temperature setpoint, the "off" (lag) boiler pump shall start. And, upon proving flow, the "off" boiler shall be enabled to fire under its factory controls to maintain 180°F. The just-started boiler's blending valve shall be controlled by an incoming 145°F water temperature low-limit sensor and setpoint similar to the lead boiler's, and subsequently, in unison with the other boiler's blending valve to maintain the reset, secondary hot-water temperature.

Anytime both boilers are operating and their control valves are less than 40% open to the secondary return line, the boiler and pump that has run longest shall shut down.

Modular Boilers

Modular boilers provide heat over a large range of loads and avoid standby and other losses associated with operating large boilers at small loads. Fig. 14 shows a primary-secondary piping arrangement where each modular boiler has its own pump. The boiler pump is on when the boiler is on.

Boilers that are off have no flow and are allowed to cool. Each boiler that is on operates at or near full capacity. Avoiding intermittent operation prevents losses up the stack or to the surrounding area when the boiler is off.

Normal control of modular boilers cycles one of the on-line boilers to maintain water temperature in the supply main to meet load requirements. The supply main's control sensor cycles the boilers in sequence. If the load increases beyond the capacity of the boilers that are on-line, an additional boiler is started. The lead (cycling) boiler can be rotated on a daily or weekly basis to equalize wear among all boilers or when using digital controls, the program can start the boiler that has been off the longest.

CONCLUSION

In many facilities, boilers represent the most significant pieces of energy consuming equipment. Understanding how a boiler functions, and how it can best be controlled may lead to large energy savings for residential, commercial, and industrial complexes.

REFERENCES

1. *Engineering Manual of Automatic Controls*, Honeywell International: Golden Valley, MN, 2005; available online at www.buildingcontrolworkbench.com.

BIBLIOGRAPHY

1. Boilers and Fired Systems. Chapter 5 in *Energy Management Handbook*, 5th Ed., Turner, W.C., Ed., Fairmont Press: Lilburn, GA, 2006.
2. Petchers, N. Boilers. *Combined Heating, Cooling, and Power Handbook*; Fairmont Press: Lilburn, GA, 2003; (Chapter 7).
3. Boilers, Chapter 27, *ASHRAE Handbook of HVAC Sytems and Equipment*; American Society of Heating, Air Conditioning, and Refrigerating Engineers: Atlanta, GA, 2004.
4. Heselton Ken, *Boiler Operators Handbook*, Fairmont Press: Lilburn, GA.
5. Harry Taplin. *Boiler Plant and Distribution System Optimization Manual*; Fairmont Press: Lilburn, GA.

Building Automation Systems (BAS): Direct Digital Control

Paul J. Allen
Reedy Creek Energy Services, Walt Disney World Co., Lake Buena Vista, Florida, U.S.A.
Rich Remke
Commercial Systems and Services, Carrier Corporation, Syracuse, New York, U.S.A.

Abstract
This chapter is designed to help energy managers understand some of the fundamental concepts of Building Automation Systems (BAS). A BAS is used to control energy consuming equipment—primarily for heating, ventilating and air conditioning (HVAC) equipment and lighting controls. We thoroughly examine each component of a BAS in today's BAS technology and what a BAS might look like in the future. The BAS of tomorrow will rely heavily on the Web, TCP/IP, high-speed data networks, and enterprise level connectivity.

INTRODUCTION

The combination of low-cost, high-performance microcomputers together with the emergence of high-capacity communication lines, networks, and the Internet has produced explosive growth in the use of Web-based technology for direct digital control (DDC) building automation systems (BAS).[1] Many of these current BAS systems use a proprietary information structure and communications protocol that greatly limits the plug-and-play application and addition of interchangeable components in the system. Control solutions such as BACnet and LonWorks have helped this situation somewhat, but they have also introduced their own levels of difficulties. The BAS of the future will integrate state-of-the-art information technology (IT) standards used widely on the Internet today. These new IT-based systems are rapidly overtaking the older BAS systems. All of the established BAS companies are quickly developing ways to interface their systems using IT standards to allow the use of Web browsers such as Internet Explorer and Netscape Navigator.

This article will examine all facets of a BAS, from field controllers to the front-end interface. The emphasis is on understanding the basic BAS components and protocols first, and then examining what a BAS might look like in the future based on the influence of IT standards. Finally, this article will discuss upgrade options for legacy BAS systems and BAS design strategies.

Even though we will be referring exclusively to the term BAS in this chapter, the building automation controls industry also uses the following terms interchangeably: direct digital control (DDC), energy management system (EMS), building automation and control system (BACS), and building management system (BMS).

Keywords: Building automation systems, BAS; Direct digital control, DDC; Internet; LonWorks; BACnet.

THE BASICS OF TODAY'S BAS

At a minimum, a BAS is used to control functions of a heating, ventilating, and air conditioning (HVAC) system, including temperature and ventilation, as well as equipment scheduling. Additional basic features include the monitoring of utility demand, energy use, building conditions, climatic data, and equipment status. Even basic BAS are generally expected to perform control functions that include demand limiting and duty cycling of equipment. Building automation systems report outputs can show the facility utility load profiles, the trends and operation logs of equipment, and the generation of maintenance schedules.

More elaborate BAS can integrate additional building systems—such as video surveillance, access control, lighting control, and interfacing—with the fire and security systems. However, in large organizations and on campuses today, it is still more common to see dedicated systems for these additional building systems due to divisions in management functional responsibility, code issues, and the features and performance of dedicated systems.

Today's BAS are expected to receive and process more sophisticated data on equipment operation and status, such as data from vibration sensors on motors, ultrasonic sensors on steam traps, infrared sensors in equipment rooms, and differential pressure sensors for filters. Top-of-the-line BAS today also have additional capabilities, such as chiller/boiler plant optimization, time schedule and setpoint management, alarm management, and tenant billing to name a few. Most BAS manufacturers today have started to offer some form of Web-based access to their existing control systems and are actively developing Web-based capability for their future products.

CONTROLLER-LEVEL HARDWARE AND SOFTWARE

Controller Hardware

Building automation systems controllers are used to provide the inputs, outputs, and global functions required to control the mechanical and electrical equipment. Most BAS manufacturers provide a variety of controllers tailored to suit the specific need. Shown below is a list of the most common BAS controllers.

Communications interface: Provides the communication interface between the operator workstation and the lower-tier controller network. On a polling controller network, a communications interface is used to transfer data between the controllers.

Primary controller: Provides global functions for the BAS control network that can include real-time clocks, trend data storage, alarms, and other higher-level programming support. Some BAS manufacturers combine all these functions into one primary controller, while other manufacturers have separate controllers that are dedicated to each global function.

Secondary controller: Contains the control logic and programs for the control application. Secondary controllers usually include some on-board input/output (I/O) and may interface to expansion modules for additional I/O. Inputs include temperatures, relative humidity, pressures, and fan and pump status. Outputs include on/off and valve or damper control. Also included in this group are application-specific controllers that have limited capability and are designed for a specific task. Examples include controllers for variable air volume terminal unit (VAV) boxes, fan coil units, and multistage cooling and heating direct-expansion (DX) air conditioning systems.

For further reference, the Iowa Energy Center has an excellent Web site (www.ddc-online.org) that shows a complete overview of the designs, installations, operation, and maintenance of most BAS on the market today.

Controller Programming

Building automation systems controllers typically contain software that can control output devices to maintain temperature, relative humidity, pressure, and flow at a desired setpoint. The software programming can also adjust equipment on/off times based on a time-of-day and day-of-week schedule to operate only when needed.

The software used to program the controllers varies by BAS manufacturer and basically falls into three categories:

1. Fill-in-the-blank programming using standard algorithms
2. Line-by-line custom programming
3. Graphical custom programming.

Fill-in-the-blank: This type of programming uses precoded software algorithms that operate in a consistent, standard way. The user fills in the algorithm configuration parameters by entering the appropriate numbers in a table. Typically, smaller control devices use this type of programming, like those that control a fan coil or VAV box controller. These devices all work the same way and have the same inputs and outputs.

A few manufacturers have used fill-in-the-blank programming for devices that are more complex with which a variety of configurations can exist, such as air handlers. Standard algorithms are consistent for each individual component. As an example, the chilled-water valve for an air-handling unit is programmed using the same standard algorithm with only the configuration parameters adjusted to customize it for the particular type of valve output and sensor inputs. Programming all of the air-handler devices using the appropriate standard algorithm makes the air-handling unit work as a system.

The advantage of fill-in-the-blank standard algorithms is that they are easy to program and are standard. The downside is that if the standard algorithm does not function as desired, or if a standard algorithm is not available, the system requires development of a custom program.

Line-by-line custom programming: Control programs are developed from scratch and are customized to the specific application using the BAS manufacturer's controls programming language. In most cases, programs can be reused for similar systems with modifications as needed to fit the particular application.

The advantage of line-by-line custom programs is that technicians can customize the programs to fit any controls application. The disadvantage is that each program is unique, and troubleshooting control problems can be tedious, because each program must be interrogated line by line.

Graphical custom programming: Building automation systems manufacturers developed this method to show the control unit programs in a flowchart style, thus making the programming tasks more consistent and easier to follow and troubleshoot.

Below are some additional issues to consider regarding control unit programming:

- Can technicians program the control units remotely (either network or modem dial-in), or must they connect directly to the control unit network at the site?
- Does the BAS manufacturer provide the programming tools needed to program the control units?
- Is training available to learn how to program the control units? How difficult is it to learn?
- How difficult is it to troubleshoot control programs for proper operation?

Controller Communications Network

The BAS controller network varies depending on the manufacturer. Several of the most common BAS controller networks used today include RS-485, Ethernet, attached resource computer network (ARCNET), and LonWorks.

RS-485. This network type was developed in 1983 by the Electronic Industries Association (EIA) and the Telecommunications Industry Association (TIA). The EIA once labeled all its standards with the prefix "RS" (recommended standard). An RS-485 network is a half-duplex, multidrop network, which means that multiple transmitters and receivers can exist on the network.

Ethernet: The Xerox Palo Alto Research Center (PARC) developed the first experimental Ethernet system in the early 1970s. Today, Ethernet is the most widely used local area network (LAN) technology. The original and most popular version of Ethernet supports a data transmission rate of 10 Mb/s. Newer versions of Ethernet called "Fast Ethernet" and "Gigabit Ethernet" support data rates of 100 Mb/s and 1 Gb/s (1000 Mb/s).

ARCNET: A company called Datapoint originally developed this as an office automation network in the late 1970s. The industry referred to this system as ARC (Attached Resource Computer) and the network that connected these resources as ARCNET. Datapoint envisioned a network with distributed computing power operating as one larger computer.

LonWorks: This network type was developed by Echelon Corporation in the 1990s. A typical node in a LonWorks control network performs a simple task. Devices such as proximity sensors, switches, motion detectors, relays, motor drives, and instruments may all be nodes on the network. Complex control algorithms, such as running a manufacturing line or automating a building, are performed through the LonWorks network.

Controller Communications Protocol

A communications protocol is a set of rules or standards governing the exchange of data between BAS controllers over a digital communications network. This section describes the most common protocols used in a BAS.

BACnet: A data communication protocol for building automation and control networks is a standard communication protocol developed by the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) specifically for the building controls industry. It defines how applications package information for communication between different BAS. The American National Standards Institute (ANSI) has adopted it as a standard (ASHRAE/ANSI 135-2001).

LonTalk: An interoperable protocol developed by Echelon Corporation and named as a standard by the Electronics Industries Alliance (ANSI/EIA-709.1-A-1999). Echelon packages LonTalk on its "neuron chip," which is embedded in control devices used in a LonWorks network.

Proprietary RS-485: The protocol implemented on the RS-485 network is usually proprietary and varies from vendor to vendor. The Carrier Comfort Network (CCN) is an example of a proprietary RS-485 communications protocol.

Modbus: In 1978, Modicon developed the Modbus protocol for industrial control systems. Modbus variations include Modbus ASCII, Modbus RTU, Intel® Modbus RTU, Modbus Plus, and Modbus/IP. Modbus protocol is the single most-supported protocol in the industrial controls environment.

TCP/IP: Transmission Control Protocol/Internet Protocol (TCP/IP) is a family of industry-standard communications protocols that allow different networks to communicate. It is the most complete and accepted enterprise networking protocol available today, and it is the communications protocol of the Internet. An important feature of TCP/IP is that it allows dissimilar computer hardware and operating systems to communicate directly.

ENTERPRISE-LEVEL HARDWARE AND SOFTWARE

Client Hardware and Software

Normally, a personal computer (PC) workstation provides operator interface into the BAS. The PC workstation may or may not connect to a LAN. If a server is part of the BAS, the PC workstation would need LAN access to the server data files and graphics. Some smaller BAS use standalone PCs that have all the BAS software and configuration data loaded on each PC. Keeping the configuration data and graphics in sync on each PC becomes problematic with this design.

A graphical user interface (GUI) is one of the client-side software applications that provide a window into the BAS. The GUI usually includes facility floor plans that link to detailed schematic representations and real-time control points of the building systems monitored by the BAS. The GUI allows technicians to change control parameters such as setpoints and time schedules, or to override equipment operations temporarily. Other client-side software applications include the following:

- Alarm monitoring
- Password administration
- System setup configuration
- Report generation
- Control-unit programming and configuration.

Server Hardware and Software

Servers provide scalability, centralized global functions, data warehousing, multiuser access, and protocol translations for a midsize to large BAS. Servers have become more prominent in the BAS architecture as the need has grown to integrate multivendor systems, publish and analyze data over an intranet or extranet, and provide multiuser access to the BAS. While having a central server on a distributed BAS may seem contradictory, in reality, a server does not take away from the standalone nature of a distributed control system. Servers enhance a distributed control system by providing functions that applications cannot perform at the controller level. In fact, a BAS may have several servers distributing tasks such as Web publishing, database storage, and control system communication.

Servers provide the ability to control a BAS globally. Facilitywide time scheduling, load shedding, and setpoint resets are examples of global functions a BAS server can perform. Because these types of functions are overrides to the standard BAS controller-level programs, having them reside in the server requires that steps be taken to ensure continued control system operation should the server go down for any length of time. The distributed BAS should have the ability to "time out" of a server override if communications with the server is lost. When the server comes back online, the BAS should have rules that govern whether the override should still be in effect, start over, or cancel. Servers also can perform computational tasks, offloading this work from the BAS control units.

BAS DESIGN ISSUES

Aside from the impact that IT will have on future EMS, there are some fundamental characteristics that owners have always desired and will continue to desire from a BAS:

- Having a single-seat user interface
- Compatible with existing BAS
- Easy to use
- Easily expandable
- Competitive and low cost
- Owner maintainable.

There have been several changes made by the BAS industry to help satisfy some of these desires. The creation of "open" protocols such as LonWorks and BACnet has made field panel interoperability plausible. The development of overlay systems that communicate with multiple BAS vendor systems has made a single-seat operation possible. However, each has introduced its own levels of difficulties and additional cost.

Users that master their BAS are more likely to be successful than users that delegate the responsibility to someone else. The BAS vendor should be a partner with the owner in the process, not the master.

New-Facility BAS Design

There are two strategies available for the design and specification of BAS for new facilities:

1. Specifying a multivendor interoperable BAS
2. Standardizing on one BAS manufacturer's system.

Specifying a multivendor interoperable BAS is probably the most popular choice among the facility design community. The engineer's controls design is more schematic and the specifications are more performance based when this approach is used. In other words, the engineer delegates the responsibility of the detailed BAS design to the temperature controls contractor because the engineer does not actually know which BAS vendor will be selected. Even if only one BAS vendor was selected, it is very rare that the engineer would be intimately knowledgeable with this system anyway. Therefore, the resulting BAS design is by nature somewhat vague and entirely performance based. The key to making this approach successful is in the details of the performance specification, which is not a trivial task. Competition results from multiple BAS vendors bidding on the entire BAS controls installation.

The second approach is based on standardizing on one BAS manufacturers system. To create competition and keep installation cost low, the engineer must create the BAS design as part of his or her design documents and prescriptively specify all components of the BAS. This allows multiple temperature control contractors to bid on the BAS installation (wire, conduit, and sensor actuators)—everything outside of the BAS field panel. Everything inside the BAS field panel is owner furnished. Contractors familiar with the owners' BAS, or the owners' own technicians, perform the controller wire termination, programming, and startup. This approach is successful when all parties work together. The design engineer must produce a good BAS design. The temperature controls contractor must install the field wire, conduit, sensors, and actuators properly. Finally, the BAS contractor must terminate and program the BAS panel correctly. A successful project is a system that integrates seamlessly with the owners' existing BAS.

Upgrading an Existing BAS

Most users already own and operate a legacy BAS that they might desire to upgrade from a standalone BAS to a network-based system.[2] The benefits of a network-based BAS appear as better standard operational practices and procedures, opportunities to share cost-savings programs and strategies, and wider access to building control processes. The key to justifying the costs associated with networking a BAS is that it can be done at a reasonable cost and is relatively simple to implement and operate.

There are three main strategies available when upgrading a BAS from a standalone system to a network-based system:

1. Remove the existing BAS and replace it with a new network-based BAS.
2. Update the existing BAS with the same manufacturer's latest network-based system.
3. Install a BAS interface product that networks with an existing BAS.

The first upgrade strategy is simply to replace the existing BAS with a newer network-based BAS that has been established as a standard within your company. The cost for this option is solely dependent on the size of the BAS that will be replaced. However, this approach might be justified if the existing BAS requires high annual maintenance costs or has become functionally obsolete.

The second upgrade strategy available is to contact the original BAS manufacturer and request a proposal for its upgrade options. Most BAS manufacturers have developed some form of Ethernet network connectivity. Typically, some additional hardware and software is required to make the system work on an Ethernet network. The cost for this might be very reasonable, or it could be very expensive. It all depends on how much change is required and on the associated hardware, software, and labor cost to make it all work.

The third upgrade strategy involves the installation of a new network-based system that is specifically designed to interface with different BAS systems. These systems typically have dedicated hardware that connects to the BAS network and software drivers that communicate with the existing BAS controllers. The new BAS interface controllers also have an Ethernet connection so they can communicate on the corporate LAN. Users view the BAS real-time data by using Web browser software on their PC. The advantage of this strategy is that a multitude of different BAS systems can be interfaced. The disadvantage is that the existing BAS software must still be used to edit or add new control programs in the existing BAS field controllers.

FUTURE TRENDS IN BAS

The future of DDC in BAS can be found on the Web. Most BAS manufacturers see the need to move their products to the Internet tremendous economies of scale and synergies can be found there. Manufacturers no longer have to create the transport mechanisms for data to flow within a building or campus they just need to make sure their equipment can utilize the network data paths already installed or designed for a facility. Likewise, with the software to display data to users, manufacturers that take advantage of presentation-layer standards such as Hypertext markup language (HTML) and Java can provide the end user a rich, graphical, and intuitive interface to their BAS using a standard Web browser.

So where do we go from here?

Faster, Better, Cheaper

Standards help contain costs by not reinventing the wheel every time a new product is developed or brought to market. While there is a risk of stagnation or at least uninspired creativity using standards, Internet standards have yet to fall into this category, due to the large consumer demand for rich content on the Internet. A BAS, even at its most extensive implementation, will use only a tiny subset of the tools available for creating content on the Internet.

When a BAS manufacturer does not have to concentrate on the transport mechanism of data or the presentation of that data, new products can be created at a lower cost and more quickly. When the user interface is a Web browser, building owners can foster competition among manufacturers because each BAS system is inherently compatible with any competitors at the presentation level. All that separates one BAS from another in a Web browser is a hyperlink.

Another area where costs will continue to fall in using Internet standards is the hardware required to transport data within a building or a campus. Off-the-shelf products such as routers, switches, hubs, and server computers make the BAS just another node of the IT infrastructure. Standard IT tools can be used to diagnose the BAS network, generate reports of BAS bandwidth on the intranet, and back up the BAS database.

Owners will reap the benefits of Internet standards through a richer user interface, more competition among BAS providers, and the ability to use their IT infrastructure to leverage the cost of transporting data within a facility.

The Enterprise

Extensible markup language (XML): XML is an Internet standard that organizes data into a predefined format for the main purpose of sharing between or within computer systems. What makes XML unique is that data tags within the XML document can be custom or created on the fly and, unlike HTML tags, are not formatted for presenting the data graphically. This makes XML a great choice for machine-to-machine (M2M) communication.

Why is M2M so important? Because the next wave of BAS products will include "hooks" into other Internet-based systems. Building automation systems have done a great job of integrating building-related components together. BACnet, LonWorks, and Modbus provide the capability of connecting disparate building components

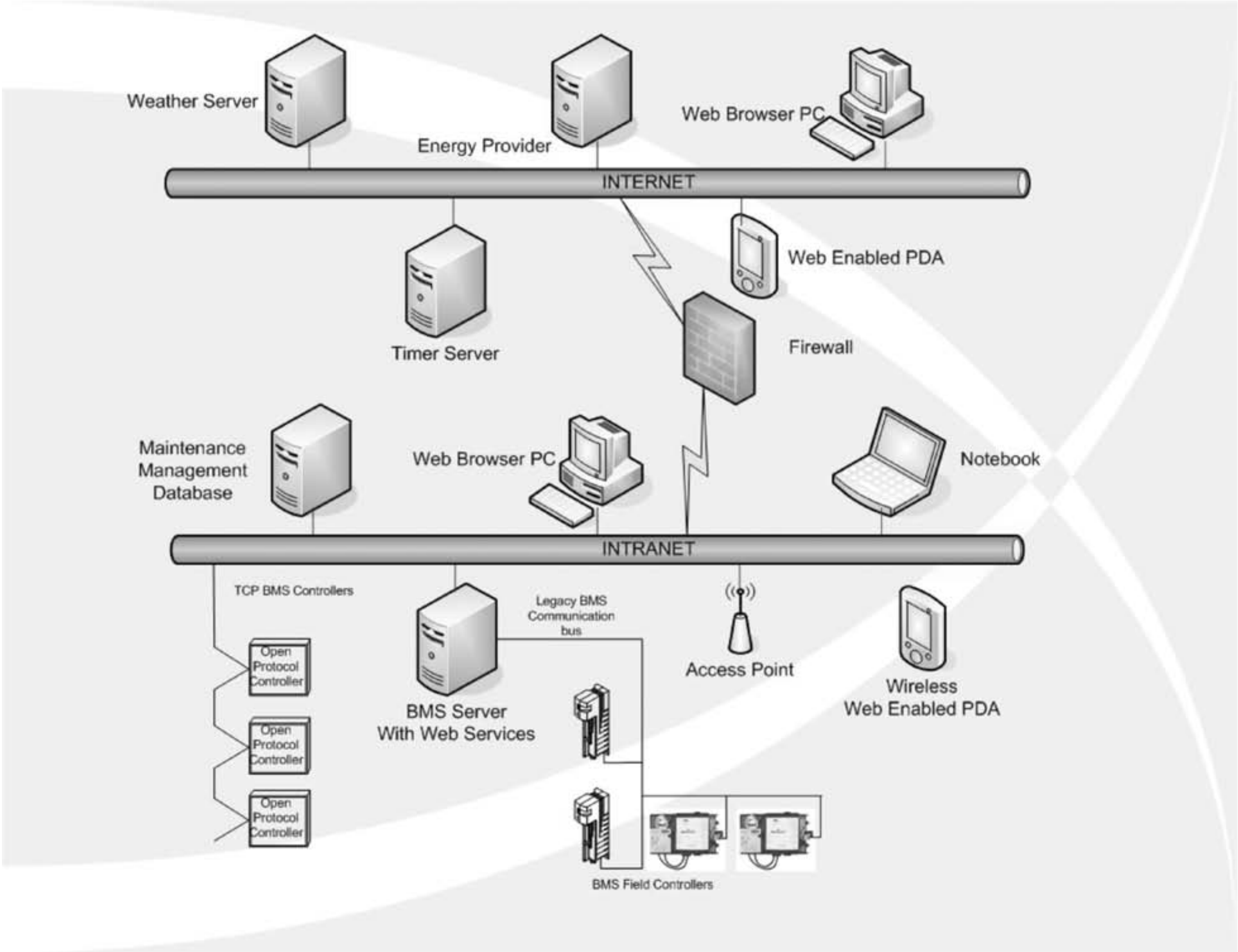


Fig. 1 Future building automation systems (BAS) network schematic.

made by different manufacturers, so that a lighting control panel can receive a photocell input from a rooftop building controller or a variable-frequency drive can communicate an alarm on the BAS when a failure occurs.

The future will require a BAS to connect to enterprise-level systems, not just building-level systems. This is where M2M and Web services come into play. Web services can be thought of as plug-ins allowing a BAS to communicate with a Web-based system or server. An example of this would be time synchronization. The Internet has many time servers that can provide the exact local time as well as greenwich mean time (GMT). A BAS can have a Web service that would plug into the BAS, synchronizing all of the time clocks within a facility with the atomic clock in Boulder, Colorado. Another example would be obtaining the outside air temperature from the local weather service. Instead of the BAS just measuring the outside air temperature at a local controller, a Web service could provide the outside air temperature, humidity, barometric pressure, and any other weather-related data. Now the BAS can make more intelligent decisions on using outdoor air for comfort cooling, determining wet-bulb setpoints for cooling towers, or even announcing that a storm is imminent.

More enticing than connecting to weather and time servers is the promise of connecting to a facility's enterprise data. The BAS of the future must become an integral part of the decision-making for allocating personnel, budgeting maintenance and upgrades, purchasing energy, and billing those that use the energy. Most larger facilities have departments that provide these types of services, yet the BAS has always stood alone, providing input through exported reports, system alarms, or human analysis. Enterprise-level integration would create Web services to connect directly to these systems, providing the data necessary for making informed decisions about capital investments, energy, or personnel. See Fig. 1 for what a BAS might look like in the future.

The good news is that XML and Web services have gained market acceptance to become the standards for enterprise-level connectivity. The bad news is that this is still in its infancy for most BAS vendors. It is a very costly effort to create an enterprise-level Web service today. Even though Web services are supported by Microsoft

Corporation, Apple Computer, Sun Microsystems, and others, they can still be custom solutions tailored to a specific accounting, maintenance management, or energy procurement system. For Web services to become mainstream in the BAS world, common services will need to be created that can be used by all BAS vendors. In addition, for Web services to be implemented properly in facilities, the skill set for BAS programmers and installers will need to include XML and a basic understanding of IP. If facility managers and technicians are to be able to make changes, adjustments, and enhancements to their enterprise system, they too will require this skill set.

The future will also need to better define the decision logic and troubleshooting tools when implementing Web services. When the BAS sends duplicate alerts to a maintenance management system, where does the logic reside to send only one technician to the trouble call? This is currently undefined. Standard tools for testing scenarios online and offline need to be developed. Even though Web services typically rely on XML, which is a self-documenting standard, XML can be very verbose. Tools need to be created to help technicians discover and correct errors quickly. When a facility decides to change its accounting system to a newer version or a different vendor, will the BAS be able to adapt? Conversion and upgrade tools need to also be considered when defining BAS Web services.

Even without all the tools identified, enterprise-level connectivity is moving ahead rapidly. The benefits of integrating BAS data within a facility's other systems can outweigh the immediate need for a complete set of tools. Web services through XML place the BAS directly into the facility data infrastructure. That's a good place to be for an energy manager wanting to maximize the investment in a facility's BAS.

CONCLUSION

The BAS of old relied heavily on a collection of separate systems that operated independently, often with proprietary communication protocols that made expansion, modification, updates, and integration with other building or plant information and control systems very cumbersome, if not impossible. Today the BAS is expected not only to handle all of the energy- and equipment-related tasks, but also to provide operating information and control interfaces to other facility systems, including the total facility or enterprise management system.

Measuring, monitoring, and maximizing energy savings are fundamental tasks of all BAS, and are the primary justification for many BAS installations. Improving facility operations in all areas, through enterprise information and control functions, is fast becoming an equally important function of the overall BAS or facility management system. The Web provides the means to share information more easily, quickly, and cheaply than ever before. There is no doubt that the Web is having a huge impact on the BAS industry. The BAS of tomorrow will rely heavily on the Web, TCP/IP, high-speed data networks, and enterprise-level connectivity. If you have not done so already, it is a good time for energy managers to get to know their IT counterparts at their facilities, along with those in the accounting and maintenance departments. The future BAS will be here sooner than you think. Get ready—and fasten your seat belts!

REFERENCES

1. Capehart, B.; Allen, P.; Remke, R.; et al. *IT Basics for Energy Managers—The Evolution of Building Automation Systems Toward the Web*; Information Technology for Energy Managers, The Fairmont Press, Inc.: Lilburn, GA, 2004.
2. Allen, P.; Remke, R.; Tom, S. *Upgrade Options for Networking Energy Management Systems*; Information Technology for Energy Managers, The Fairmont Press, Inc.: Lilburn, GA, 2005; Vol. 2.

BIBLIOGRAPHY

1. RS-485, www.engineerbob.com/articles/rs485.pdf
2. Ethernet, www.techfest.com/networking/lan/ethernet.htm
3. ARCNET, www.arcnet.com
4. LonWorks and LonTalk, www.echelon.com/products/Core/default.htm
5. BACnet, www.bacnet.org
6. Modbus, www.modbus.org/default.htm
7. XML, www.xml.com
8. Iowa Energy Office, www.ddc-online.org

Building Geometry: Energy Use Effect☆

Geoffrey J. Gilg
Pepco Energy Services, Inc., Arlington, Virginia, U.S.A.

Francisco L. Valentine
Premier Energy Services, LLC, Columbia, Maryland, U.S.A.

Abstract
Energy service companies (ESCOs) use the energy use index (EUI) as a tool to evaluate a building's potential for reduction in energy use. Select Energy Services, Inc. (SESI) has found that consideration of building geometry is useful in evaluating a building's potential for energy use reduction. Building load and energy-use simulations using Trace® and PowerDOE®, respectively, were conducted to gain insight into how building geometry impacts heating, ventilation, and air-conditioning (HVAC) sizing and energy use. The ratio of gross wall area to gross floor area, A_{wall}/A_{floor}, has been found to be a useful factor to consider when making EUI comparisons. Simulations suggest that buildings with higher A_{wall}/A_{floor} ratios require higher central plant capacities and use more energy per unit area to satisfy the heating and cooling loads. Taking a building's geometry (A_{wall}/A_{floor}) into account while estimating savings potential may produce more accurate results.

INTRODUCTION

Select Energy Services, Inc. (SESI) has conducted a multitude of building evaluations in the course of its performance contracting and design work. Select Energy Services, Inc. has many energy engineers with real-world heating, ventilation, and air-conditioning (HVAC) design experience, which often provides insight into peculiarities. One such peculiarity is "Why do two buildings of the same usage and square footage exhibit energy use indexes (EUIs) significantly different from one another?" In an attempt to answer this question, SESI conducted a series of simulations which focused on building geometry and its contribution to heating and cooling loads and annual energy use.

The tools used in this analysis are Trace® and PowerDOE®. Trace® is a software package published by C.D.S. Software that is used to determine equipment loads. PowerDOE® is published by the Electric Power Research Institute, Inc. as a "front-end" for the U.S. Department of Energy's DOE-2 building energy simulator. PowerDOE® is used to simulate annual building energy use. Both software packages allow easy and economical means to evaluate a building's HVAC capacity requirements and resulting energy use.

☆ This entry originally appeared as "The Effect of Building Geometry on Energy Use" in *Energy Engineering*, Vol. 100, No. 3, 2004. Reprinted with permission from AEE/Fairmont Press.

Keywords: Building geometry; Energy use index; Performance contracting; Energy simulation; Load calculations; Energy modeling; HVAC; Energy service company.

ENERGY USE INDEX

Even before setting foot on site, an energy service company (ESCO) can get a preliminary estimate of the potential for energy cost reduction. This can be done by analyzing the fuel and electric rates, looking for credits or rate restructuring, and evaluating EUIs. Energy use index is defined as the ratio of total annual energy used, in kBtus, divided by the square footage of the building.

$$\text{EUI} = \frac{\text{kBtu}}{\text{ft}^2} \tag{1}$$

The EUI is used as a barometer for estimating the potential for energy savings. However, it must be applied with discretion or an ESCO could pass on a great opportunity or overestimate the potential for energy cost savings.

How the Energy Use Index is Used

Once utility billing data, equipment data, and building square footage have been provided, an EUI evaluation can be conducted. The calculated EUI of the building is compared to an "ideal" EUI. The difference between the building EUI and the ideal EUI is the potential for energy savings (see Fig. 1). However, it is often cost prohibitive to attain the entire EUI differential, so ESCOs often prescribe a maximum, economically attainable, EUI improvement. Fifty percent is often used, but this depends on many factors.

From the above methodology, it is easy to see how comparing EUIs, based solely on square footage, can sometimes result in an inaccurate evaluation of the energy savings potential.

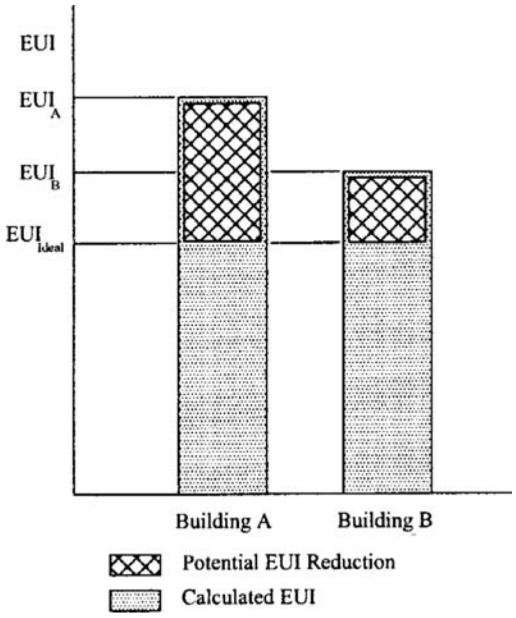


Fig. 1 Energy use comparison.

BUILDING GEOMETRY

Building geometry is an important factor to consider from a design standpoint. It influences heat loss, heat gains, infiltration, and solar gains which influence the heating and cooling load. Typically, the more wall (including windows) area available, the higher the heating and cooling loads. The first 10–15 ft from the exterior wall is considered the perimeter zone. The perimeter zone heating/cooling load is constantly changing because it is under the influence of the weather via the building envelope (walls, windows, and roof) as well as internal loads (occupants, lights, equipment, etc.). Inside this area is the interior zone, which experiences much less heating/cooling load variation (only internal loads). Therefore, if evaluating two buildings of equal floor area, use, occupancy, etc. the building with the larger interior zone will typically require less heating and cooling capacity and use less energy annually.

Building geometry is also an important factor to consider from an energy-use standpoint. If two buildings

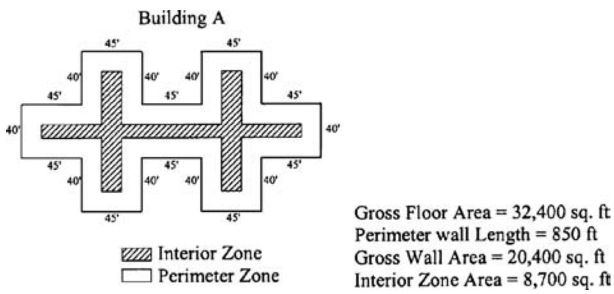


Fig. 2 Building A floor plan.

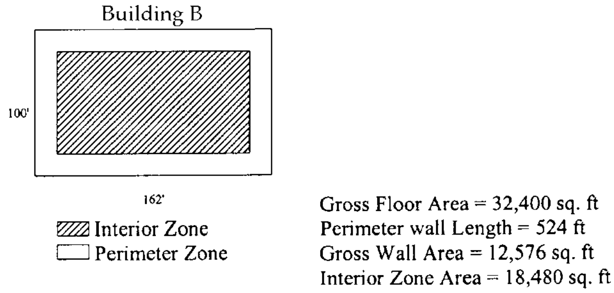


Fig. 3 Building B floor plan.

of same square footage, use, schedule, controls, occupancy, and construction exhibit significantly different EUIs, differences in building geometries may explain why. Because of its greater exposure to environmental conditions, the building with more wall area will likely have the higher EUI. Therefore, it is not uncommon to find that buildings with multiple floors or eccentric shapes use more energy than single-floor, rectangular buildings of the same square footage. For example, Figs. 2 and 3 illustrate two-story buildings, each with 32,400 ft^2 of floor space. Building A has a layout found on many military installations and educational campuses, and its protruding wings resemble radiator fins, both in cross-section and in thermal effect.

Building B is another building of the same square footage, but in the shape of a rectangle, a more compact shape with less exposed surface area.

Given the same building load parameters, Trace® calculations indicate that Building A requires 15% additional cooling capacity and 25% more heating capacity than Building B. PowerDOE® energy use simulations indicate that Building A will use 15% more energy annually than Building B. This example shows that building geometry is indeed an important factor to consider when estimating the potential for energy use reduction based on EUI comparisons. With the realization that building geometry affects energy use, how can it be accounted for, during an EUI comparison?

BUILDING MODELS

To evaluate the effect of geometry on energy use, several hypothetical models were defined, constructed, and analyzed using Trace® and PowerDOE®. The model definitions include such parameters as use (school, office, warehouse), geographical location, schedule, overall heat transfer coefficients (wall, roof, windows), and others. This methodology provides the ability to evaluate the effect of geometry in buildings with distinctly different uses and geometries. Models of the buildings were constructed in Trace® and PowerDOE® using the same building parameters. These programs allow parameters such as orientation, location, and geometry to be changed with a keystroke. A schedule of model assumptions is provided in Appendix A.

RESULTS

After conducting the building load and energy use simulations, a factor has emerged that explains why a building uses more or less energy per square foot than another. This factor takes into account differences in building geometry when evaluating energy use reduction potential. This geometric ratio (GR) is defined as the ratio of gross perimeter wall area (A_{wall}) to gross floor area (A_{floor}).

$$\text{GR} = \frac{\sum A_{\text{Wall}}}{\sum A_{\text{Floor}}} = \frac{A_{\text{Wall}}}{A_{\text{Floor}}} \qquad (2)$$

Comparisons of Trace® load calculations indicate that buildings with higher wall to floor area ratios require larger heating and cooling plants. The effect of building geometry has been found to be more pronounced as outdoor air requirements decrease. For example, Trace® load calculations (heating only) for a warehouse indicate that with each percentage point increase in the GR, the peak heating requirement increases approximately 1%. Geometry, in this example, has such a significant effect because envelope load is a larger percentage of the total heating plant requirement. Annual energy use predicted by PowerDOE®, on the other hand, is relatively flat for warehouse structures while showing significant geometric effects for school and office-type occupancies. As the outdoor air requirements increase, the contribution of envelope loads to the total heating and cooling load decreases.

Simulations have also shown that it is important to consider the percentage window area per unit wall area. The amount of window in a wall has a significant impact in the overall heat loss/gain of that exposure. Buildings with higher window to wall area ratios typically require larger heating and cooling capacities and use more energy annually per unit floor area than buildings with lower window to wall area ratios.

Trace® and PowerDOE® simulations also indicate that the orientation of a building is also an important factor to consider. If a building has a high (2 or higher) aspect (length/width) ratio and is oriented so that the long sides of the buildings are facing east–west, this building typically requires larger heating and cooling plant capacities and will consume more energy annually. Had this same building been oriented such that a long side was facing north–south, the heating and cooling plant capacity and energy use could have been reduced. Orientation of the long side of the building in a north–south direction also could have permitted more effective use of natural light to reduce lighting energy requirements which can further reduce the cooling load. Therefore, when considering the potential for energy reduction in this particular building, it is advisable to take the building's orientation into account.

Differences in building energy use can also be explained by considering geometry as it relates to the original intended use of the building. On many military installations, especially those associated with airfields, there are numerous single level, high ceiling, marginally insulated buildings whose original intended use and design are not consistent with their present utilization. For example, storage buildings and aircraft hangars are often converted to office space (without upgrading the walls, windows, or roof insulation). As such, they have undergone numerous HVAC retrofits through the years as the hangar/storage space is further converted to office space. Buildings of this type typically use more energy than buildings whose original intended use was that of an office.

The age of a building, in conjunction with geometry, also helps explain differences in EUI. Older buildings have experienced much more wear and tear and typically have higher infiltration rates. Additionally, older buildings are typically constructed with lower *R*-value materials than contemporary construction. Heating, ventilation, and air-conditioning equipment is typically at, near, or far past its useful life and requires frequent maintenance. In addition, older buildings typically are not insulated as well. Due to space restrictions, many older buildings have rooms or wings that have been added to the original building. This addition can significantly increase the perimeter wall area with only a small increase in square footage. Finally, older buildings do not benefit from recent quality control and construction standards. As a result, it is not uncommon for older buildings to exhibit higher EUIs than similar buildings of recent construction. PowerDOE® modeling has shown that building geometry, as indicated by the GR, has more influence on energy use in older buildings.

How the Geometric Ratio can be Used

When evaluating the potential for energy use reduction, it may be to the ESCO's advantage to take into account the GR of the buildings under consideration. The GR represents the influence of building geometry on energy use and can be used to gauge the effectiveness of certain energy conservation measures (ECMs). For example, ECMs associated with walls and windows such as window film, window replacements, and wall insulation upgrades may have a larger EUI impact in buildings with higher GRs. This is because wall and window conduction and solar gains are a higher percentage of the total HVAC load. Alternately, ECMs such as air-side economizers, lighting retrofits, and roof insulation upgrades may have a larger EUI impact in building with lower GRs. This is because the percentage of contribution of outdoor air, lighting, and roof conduction to the total building load is typically higher for buildings with lower GRs.

CONCLUSIONS

At the conclusion of the calculations and simulations using Trace® and PowerDOE®, SESI believes that it has, at least in small part, contributed to a better understanding of building geometry and its impact on heating and cooling requirements and energy use. This understanding can be used to better estimate the effectiveness of certain ECMs, which can help to avoid underestimation as well as overestimation of potential EUI improvement. This exercise has shown that there is value in considering building geometry while estimating the potential for energy use reduction.

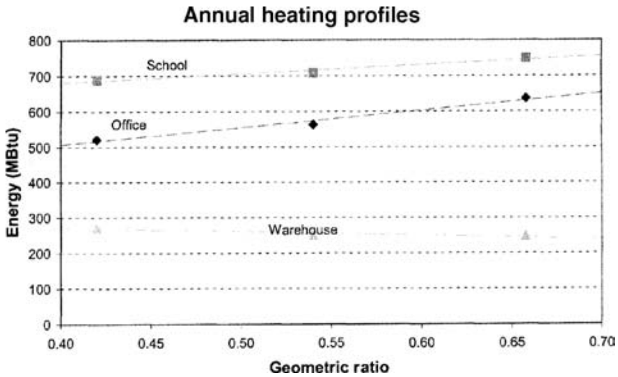
APPENDIX A. MODEL ASSUMPTIONS

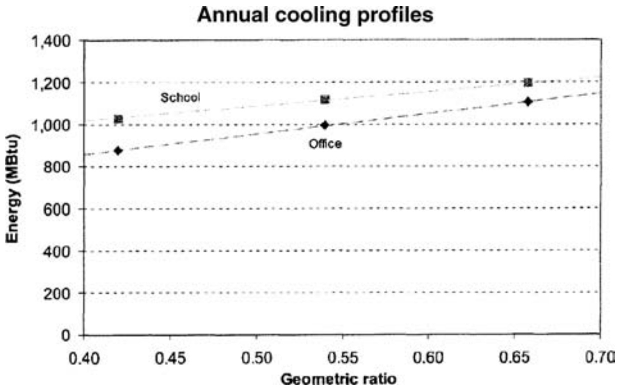
Common assumptions	
Location	Washington, DC
System type	VAV w/hot water reheat coils (VAV=Variable Air Volume)
Heating thermostat setpoint (°F)	68
Cooling thermostat setpoint (°F)	78
Floor to floor (ft)	12
Plenum height (ft)	3
Floor type	Slab on grade (6 concrete and 12 soil)
Roof type	Flat—built up
Infiltration rate (ft^3/ft^2 wall)	0.15 ft^3/ft^2 wall area
Sensible load (Btu/h per person)	250
Latent load (Btu/h per person)	200
U_{wall} (Btu/(h ft^2))	0.1 Btu/(h ft^2)
U_{window} (Btu/(h ft^2))	0.5 Btu/(h ft^2)
U_{roof} (Btu/(h ft^2))	0.05 Btu/(h ft^2)
Schedule	100% (weekdays, 8 A.M.–5 P.M.)
	20% (weekdays, 5 P.M.–8 A.M.)
	20% (weekends and holidays)
Ignore	Gymnasium, library, locker rooms, rest rooms, cafeteria, lounges, corridor spaces

Schools			
Parameter	Model #1	Model #2	Model #3
Length × width (ft)	200 × 80	100 × 80	73 × 73
Number of floors	1	2	3
Orientation (long side)	15 South of due east	15 South of due east	15 South of due east
System type	VAV w/hot water reheat	VAV w/hot water reheat	VAV w/hot water reheat
Window percentage (of wall area)	20	20	20
Occupancy (people/1000 ft^2)	50/ 1000 ft^2	50/ 1000 ft^2	50/ 1000 ft^2
Ventilation (ft^3/min/person)	15	15	15
Lighting load (W/ft^2)	1.5	1.5	1.5
Equipment load (W/ft^2)	0.25	0.25	0.25
Offices			
Parameter	Model #1	Model #2	Model #3
Length × width (ft)	200 × 80	100 × 80	73 × 73
Number of floors	1	2	3
Orientation (long side)	15 South of due east	15 South of due east	15 South of due east
System type	VAV w/hot water reheat	VAV w/hot water reheat	VAV w/hot water reheat
Window percentage (of wall area)	20	20	20
Occupancy (people/1000 ft^2)	7	7	7
Ventilation (ft^3/min/person)	20	20	20
Lighting load (W/ft^2)	1.5	1.5	1.5
Equipment load (W/ft^2)	0.5	0.5	0.5
Warehouse			
Parameter	Model #1	Model #2	Model #3
Length × width (ft)	200 × 80	100 × 80	73 × 73
Number of floors	1	2	3
Orientation (long side)	15 South of due east	15 South of due east	15 South of due east

System type	Gas unit heaters	Gas unit heaters	Gas unit heaters
Window percentage (of wall area)	20	20	20
Occupancy (people/1000 ft^2)	5	5	5
Ventilation ($ft^3/min/ft^2$)	0.05	0.05	0.05
Lighting load (W/ft^2)	1.5 W	1.5	1.5
Equipment load (W/ft^2)	0.25 W	0.25	0.25
Examples			
Parameter	Model #1-Building A	Model #2-Building B	Model #2
Length × width (ft)	See Fig. 2	See Fig. 3	100 × 80
Number of floors	3	3	2
Orientation (long side)	15 South of due east	15 South of due east	75

System type	VAV w/hot water reheat	VAV w/hot water reheat	VAV w/hot water reheat
Window percentage (of wall area)	20	20	20
Occupancy (people/1000 ft^2)	50	50	50
Ventilation (ft^3/min/person)	20	20	20
Lighting load (W/ft^2)	1.5	1.5	1.5
Equipment load (W/ft^2)	0.5	0.5	0.5





BIBLIOGRAPHY

1. ASHRAE-Fundamentals, American Society of Heating, Refrigeration, and Air-Conditioning Engineers, Inc., 2001.
2. Trane Air Conditioning Manual, The Trane Company, 1996.

Building System Simulation

Essam Omar Assem[1]
Arab Fund for Economic and Social Development, Arab Organizations Headquarters Building, Shuwaikh, Kuwait

Bene—Clim

Abstract
This entry addresses the topic of building energy system simulation and touches on its importance, benefits, and relevance to the research, architecture, and engineering communities. The entry first gives an illustration of the different energy flow paths that occur in buildings and then describes the main features of the approaches that have evolved over the past four decades in an attempt to reach a solution to those energy flow paths. In addition, an overview of the simulation process in general terms is given, with certain simulation output parameters highlighted. The entry closes with real-life examples to give the reader a sense of some of the important applications of building energy system simulation programs.

INTRODUCTION

In the context of energy systems, the term building system simulation (BSS) refers to the application of computer programs designed specifically to simulate the energy systems that exist within buildings so that a realistic prediction of the thermal performance of a building can be achieved. The most common thermal performance parameters that are referred to by heating, ventilation, and air conditioning (HVAC) engineers are the peak cooling/heating load and energy consumption. The former can be used to determine the size of the HVAC system, whereas the latter can be used to determine the system's energy efficiency. The main advantage of using BSS programs is that the thermal performance of the energy systems in a building can be determined and possibly optimized during the design stage.

For any building, the main physical issues that need to be addressed by the BSS are shown in Fig. 1 in terms of energy flow paths, which can be summarized as follows:

- Surface convection represents the heat flow from an opaque surface to the adjacent air. This process occurs on both the internal and external surfaces of a building wall. The external convection process is influenced by the surface finish, wind speed, and direction, whereas the internal convection process is influenced by the forced air flow injected by the mechanical system serving the building zone and, in the absence of a mechanical system, is influenced by the buoyancy-driven natural convection.
- Inter-surface longwave radiation represents the amount of radiation, which is a function of surface temperatures, emitted from one surface to other visible surfaces, and is influenced by surface emissivity and geometry.
- The intensity of shortwave radiation on a surface is a function of time, surface geometry, and position of the sun relative to the geographical location of the building. The magnitude of shortwave radiation penetrating an opaque surface is influenced by the surface absorptivity and, in the case of transparent surfaces (i.e., glass), is influenced by surface absorptivity, transmissivity, and reflectivity.
- Shading caused by external obstructions to the sun-ray path affects the amount of solar radiation intensity on the external surface of an opaque element and the amount of shortwave radiation penetrating a transparent element, such as a window with overhang. In addition, the amount of solar radiation penetrating a transparent surface and the internal surfaces receiving this radiation are a function of the angle of incidence of the solar radiation and the geometry of the transparent surface, as well as the geometry of the internal surfaces.
- The process of air flow within buildings is affected by the building resistance to the unidirectional air leakage between outside to inside, by the amount of air being circulated among building zones, and also by air circulation within each zone. The amount of air leakage between the outside and the inside of the building is influenced by the external air boundary conditions (such as wind speed, pressure, and air temperature), and by the internal air temperature and pressure.
- There is also the process of casual heat gain from lighting fixtures, equipment, and people. This heat gain will constitute both radiative and convective components.

Keywords: Building energy simulation; Peak load; Annual energy consumption; Thermal comfort; Day lighting; HVAC Simulation.

[1] This entry does not represent the nature of work conducted at the Arab Fund; it was prepared by the author based on his academic and previous work experience.

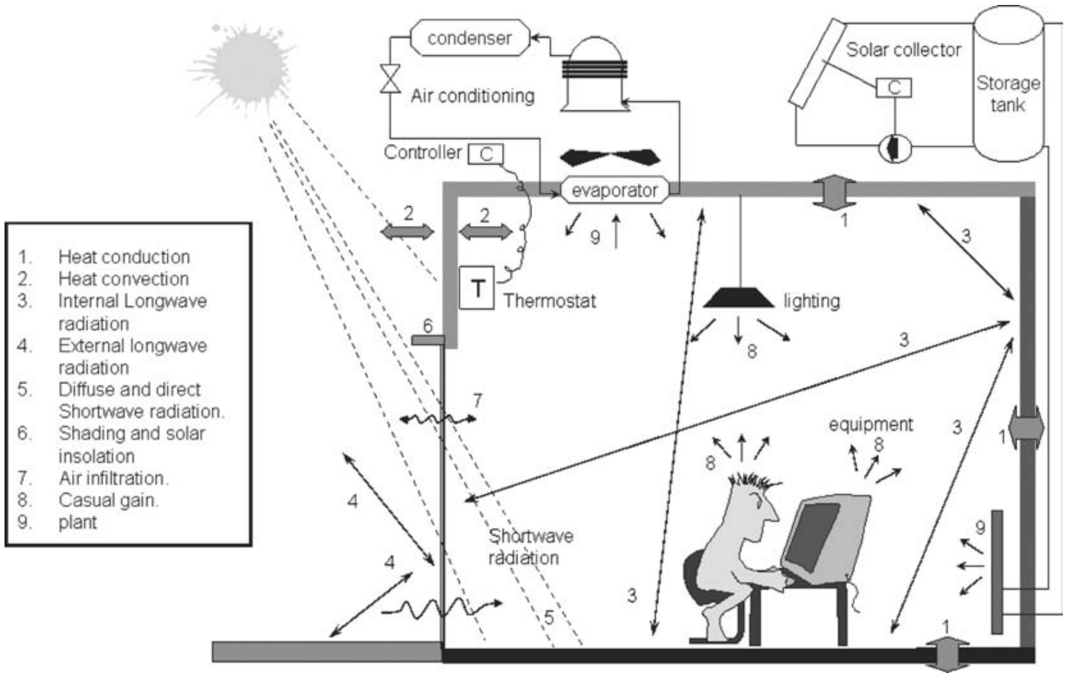


Fig. 1 Energy flow paths in buildings.

- The process of plant interaction with building zones affects their thermal comfort and can be caused by the action of a thermostat sensing and monitoring the zone's air condition. The thermostat setting can be influenced by the building occupants or by the action of an automatic control mechanism.
- Control processes are caused by the response of a plant system to dynamically changing internal conditions. The plant response is influenced by the stream of signals constantly being fed back to the various controllers that actuate the appropriate control variable according to some control action.
- Moisture transfer processes are caused not only by internal generation processes and air migration from outside, but also by the condition of the air being injected into the zone by an air conditioning system.
- In buildings, it is also common to find movable features that can be used to control the effect of certain parameters on the indoor air, such as solar radiation and ambient air. Such features are influenced by social factors and by occupant comfort level.

HISTORICAL DEVELOPMENT

From the early 1960s until the late 1980s, many BSS programs have been developed to tackle the energy flow paths described earlier. These ranged from manual methods, in which the complexity of the real system is reduced to lessen the computational overhead and the input demands on the user, to highly sophisticated approaches in which mathematical models are constructed to represent the interactions observed in reality as closely as possible.

The evolution of design tools,[3] from the traditional to the present-day simulation approach, is summarized in Table 1. First-generation models focused on a simple "handbook" approach, which is piecemeal in that no coupling between the various discrete calculations was made. A steady-state U-value calculation may be conducted to evaluate envelope heat loss, for example; then a lookup table may be used to determine an allowance for zone solar gain; and finally, a degree-day relationship may be used to predict long-term energy requirements. In such calculations, many assumptions are encompassed to simplify their application. In addition, the calculations are based on steady-state, unidirectional wall heat flow, which rarely occurs in the real world.

Second-generation models took into account the temporal aspect of energy flow, especially in the case of elements with large time constants, such as in multilayered constructions. Analytical methods such as the response factor,[10] time, or frequency domain were applied based on the assumption that the building system is linear and possesses heat transfer properties that are time invariant. Modeling integrity remained low because of the decoupling and simplifying assumptions applied to the flow paths. Although the problem encountered with

Table 1 Energy models evolution

Evolution stage	Main features	
1st Generation (traditional)	Handbook orientated	Indicative
	Simplified	Application limited
	Piecemeal	Difficult to use
2nd Generation	Dynamics important	↓
	Less simplified	
	Still piecemeal	
3rd Generation (current)	Field problem approach	
	Move to numerical methods	
	Integrated view of energy heat and mass transfer considered	
	Better user interface	
	Partial computer aided building design (CABD) integration	
4th Generation (next)	CABD integration	
	Advanced numerical methods	Predictive
	Intelligent knowledge based	Generalized
	Advanced software engineering	Easy to use

first-generation models related to quantifying transient heat conduction through multilayered constructions was overcome by second-generation models, a number of new problems emerged. These problems include the need for fast, powerful computers; an extensive input data set to define building geometry; and more appropriate user interfaces.

In recent years, numerical methods in third-generation models have played an increasingly important role in the analysis of heat transfer problems. Numerical methods can be used to solve complex, time-varying problems of high and low order. All fundamental properties are assumed to be time dependent, and coupling among the temporal processes is accounted for. The method allows for high-integrity modeling. Differential equation sets, representing the dynamic energy and mass balances within combined building and plant networks, are solved simultaneously and repeatedly at each (perhaps variable) time step as a simulation is conducted under the influence of control action. Modeling integrity has been increased so that the system is more representative of reality and is easy to use because of the improved graphical I/O (input/output) user interface. The highly transient nature and dynamic response of HVAC equipment are now fully taken into account.

TYPICAL APPLICATIONS

Building system simulation programs can be applied in many situations; they are used by researchers, engineers, and architects to solve real-life problems. In addition to the calculation of the peak cooling/heating loads and annual energy consumption, the following examples give a sense of other situations in which BSS can be a useful tool:

- Predicting daylight levels within building spaces to minimize the energy consumed by artificial lighting and to improve visual comfort and productivity.
- Predicting the pattern of air distribution within spaces to optimize the air conditioning duct design.
- Predicting the effect of using innovative building envelope designs (such as vacuum-insulated wall panels and dual-skin facades) on the overall performance of buildings and on thermal comfort.
- Given a number of known energy conservation measures, BSS can be used to identify the most cost-effective option so that the capital investments are justified.
- For existing buildings, computer models can be developed and calibrated to determine the energy savings corresponding to alternative energy conservation measures or to alternative building and HVAC systems operational strategies.
- The impact of renewable technologies, such as solar collectors and photovoltaic, on energy consumption can be investigated.
- Application of passive heating/cooling concepts can be predicted easily by using simulation programs, and alterations to the building design can be made accordingly.
- The dynamic response of buildings when subjected to temporal boundary conditions can be investigated.

- Building system simulation are used extensively to develop codes and standards for energy conservation in different classes of buildings.

THE SIMULATION PROCESS

In any BSS program, the whole process can be represented by three main stages: model definition, simulation, and result analysis and interpretation, as indicated in Fig. 2. Depending on the level of detail, the building first has to be described in a manner that is acceptable to the program. The early versions of the DOE-2,[5] for example, required the use of a special description language to define the input parameters associated with the building under consideration. This required the user to learn the language in addition to learning the use of the simulation and output data analysis. To simplify the model definition process, graphical user interfaces (GUI) were later developed so that the errors at the input stage can be minimized. In any way, the input data describing the building and its systems then are fed to the solver (simulator) to perform the simulation according to a predefined schedule in which the simulation period and output parameters are specified. When the output is obtained from a simulation, it can be checked, and if necessary, the input stage may be repeated to rectify any errors made and to repeat the simulation. Even with the sophisticated GUIs available with many BSS programs, a verification procedure must be followed to ensure that the results of the simulation are reasonable and acceptable.

The most common input parameters required to simulate a building system are

- *The weather data*. Usually, this represents hourly data for many parameters, such as the dry bulb temperature; relative humidity; wind speed and direction; and diffuse, direct, and global solar radiation. A typical metrological year can be established to represent the weather pattern for a particular location, using at least 10 years' worth of measured weather data.
- *The building location*. This can be established from the geographical location of the building site in terms of longitude and latitude.
- *The building geometry*. This is determined by defining each external surface's geometry and orientation (walls and roofs).
- *Building materials*. The construction materials used in walls and roofs are defined in terms of layers, including the thermophysical properties such as the density, specific heat capacity, and thermal conductivity for each layer.
- *Casual loads*. Those are internal loads from lighting, appliances, and people, which are specified with a predefined schedule.

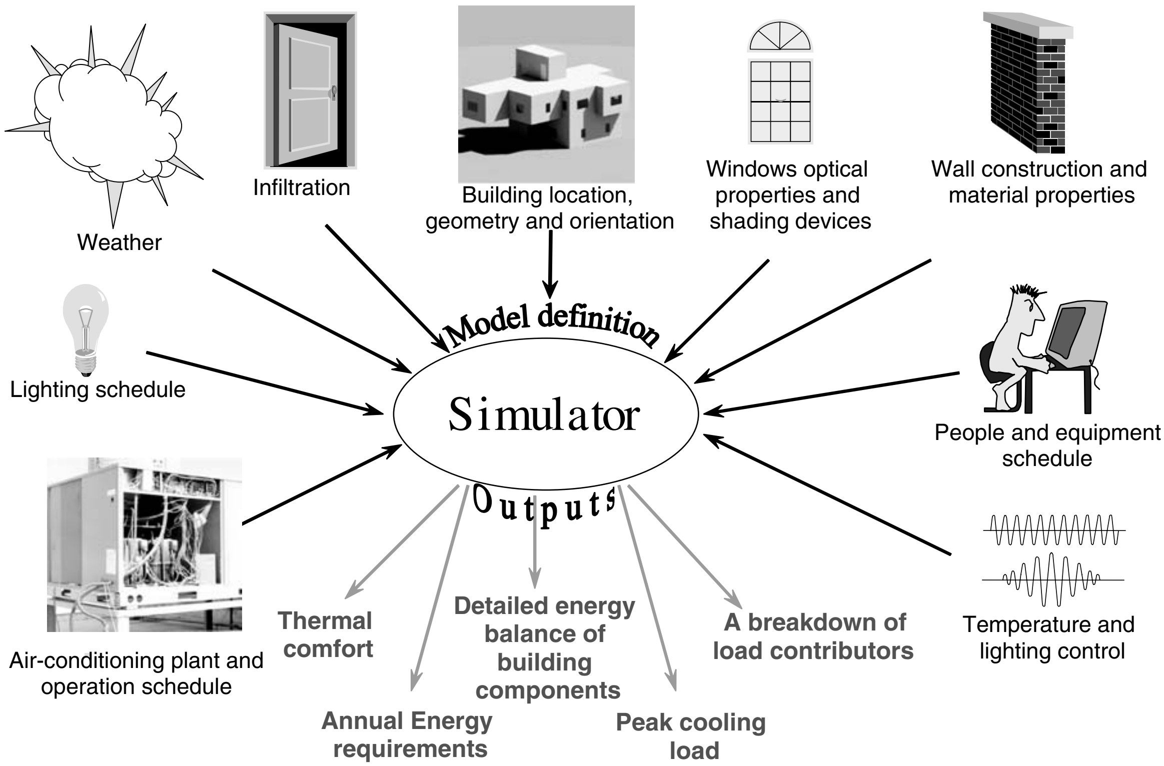


Fig. 2 The simulation process in terms of inputs and output.

- *Shading devices*. They could be used on an opaque surface (walls or roof) and transparent surfaces (windows and glass doors).
- *Transparent surfaces*. Their size, geometry, and optical properties (visible transmittance, solar transmittance, absorptance, and U-value) for both exterior and interior surfaces.
- *Plant*. This is the system used to maintain the internal environment of a building at a certain temperature and humidity, such as the HVAC system.
- *Infiltration*. This relates to the uncontrolled bidirectional flow between the inside of the building and the outside.
- *Control*. The control strategy adopted by the HVAC system can be defined by specifying the properties of the controller and the range of conditions to perform the control. In addition, lighting controls can be specified for situations in which daylight utilization may be required.

The input parameters mentioned above are adequate in most cases to perform the simulation. With some state-of-the-art BSS programs such as the ESP-r,[6] however, it is possible to define additional input data to allow addressing fairly complex issues if deemed necessary. It may be required, for example, to perform three-dimensional conduction analysis of a particular section within a building's construction to estimate the impact of thermal bridging on the heat flow through the building's envelope, which will require the definition of the domain in three-dimensional form.[2] In addition, a more accurate assessment of the infiltration levels and air movement between the building spaces may be necessary to arrive at a more realistic assessment of the thermal performance of the building, in which case an air flow network will need to be specified,[8] or if this was not adequate, a computational fluid dynamics calculation[12] can be incorporated to reach an even higher accuracy level. Furthermore, the effect of environmental systems dynamics can be investigated in detail using BSS programs such as EnergyPlus,[4] ESP-r, HVACSIM+,[9] and TRNSYS.[13] Such BSS programs were designed to study the transient effect of both the HVAC systems and the building, and with some programs, it is even possible to use smaller simulation time steps for the HVAC simulation because the HVAC systems exhibit time constants that are much smaller than the building structure.

When the input parameters are defined, the simulator performs the calculations, based on the adopted mathematical approach, to generate the output required by the user. It is possible to perform the analysis at the whole-system level (i.e., the building and plant systems) or at the system level (the building or plant), or down to the component level (i.e., a cooling coil, a fan, a window, a wall, etc.). Some of the outputs that can be obtained are shown in Fig. 2. These outputs include:

- *Thermal comfort*. This is at the whole-system level and is a very important parameter for many buildings. The simulation output can indicate whether the space in question has an acceptable thermal comfort condition for the occupants.
- *Annual energy requirements*. This is also at the whole-system level, which gives an indication of the total electrical energy consumed by the HVAC system (another source of energy, such as gas or oil, can also be included) and by the lighting and equipment or appliances used in the building.
- *Detailed energy balance of building components*. This is at the component level, for which the output data can be studied to investigate the thermal performance of individual building components such as walls, roof, and windows.
- *Peak load requirements*. This information is important because it is used to size the HVAC systems for a building. It is common to perform the calculation on the building system first to determine the peak cooling/heating load required and then to utilize the output to perform a whole-plant and building analysis to determine the peak electrical demand for the whole plant and building system.
- *Breakdown of load contributors*. This information is useful to identify the major load users so that countermeasures of more significant impact on energy use may be introduced.

To realize the benefits of BSS programs in real-life situations, two examples are discussed in the following sections. The examples focus mainly on how the problem was addressed and also on how to utilize the simulation output so that the set objectives can be achieved. The detailed input data fed to the BSS program to conduct the simulations in the examples were not addressed, because the data are fairly detailed and will not serve the purpose.

EXAMPLE 1

The dayroom area of a hospital in Scotland suffers from high temperature during summer. The dayroom area is joined to the dining-room area and consists of windows with large glazing area. The occupants of the dayroom are elderly patients, who are not always capable of judging thermal stress and so would be unlikely to leave the dayroom when overheating occurs or to take corrective action by, for example, opening a window. The objective of this study was to advise on possible modifications to the hospital ward to better control its radiation and temperature environment.[7]

A base case test was conducted for a typical Glasgow summer weather pattern for the period of Wednesday, July 7 to Thursday, July 8. In this test, it was assumed that all the

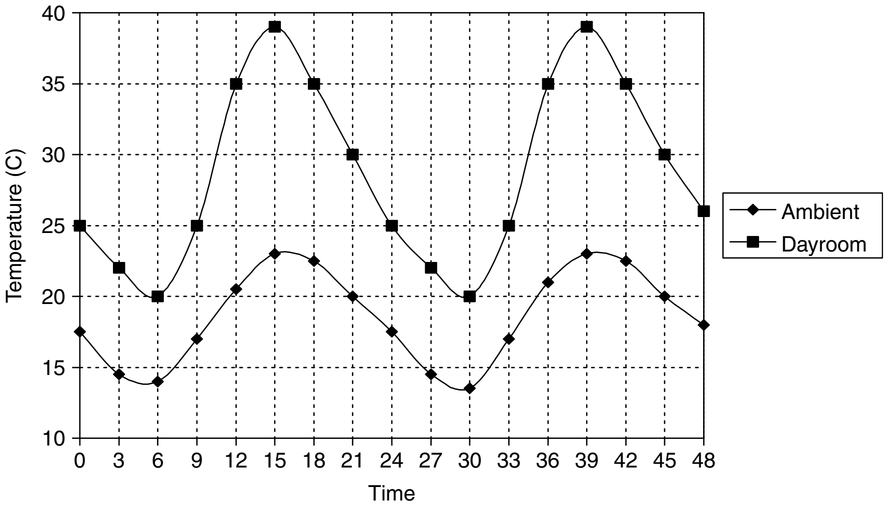


Fig. 3 Base case with no blinds and one air change per hour.

blinds were open in the dayroom and that a fixed infiltration rate of one air change per hour prevailed. Shading analysis was performed to take into account any obstructions and overhangs. The resulting temperatures are shown in Fig. 3. As indicated, the thermal stresses in the dayroom are significant. In reality, it is unlikely that fixed infiltration will occur, because this will depend on many factors, such as the magnitude of open windows, cracks, vents, and so on, as well as changes in pressures and temperatures across the building envelope. For this reason, an air flow network was set up in which the dayroom was represented by nodes on two vertical levels to account for temperature stratification; other air nodes were added to represent wind-induced pressures on the various facades. Then the air flow network nodes were connected by flow components to represent cracks, windows, and doors.

Initially, it was assumed that windows in the dayroom would remain open at all times. The effect of opening and closing of windows in the dining room, however, was considered subsequently by the addition of components to represent flow control. The control action taken was to open a window when the adjacent indoor air node temperature exceeded 20°C. This action was applied to the dining room. An extract fan component was also defined and was set to be thermostatically controlled so that air was extracted from the dayroom's upper point (hottest level) if the upper-level temperature exceeded 25°C, and the hot air was exhausted through the west wall.

A plant was also set up to represent the heating system, which is expected to be operating when room temperatures are below the desired heating set point temperature of 21°C. The drop in temperature below the heating set point occurs because it was assumed that dayroom windows are open all the time and, thus, ambient air free cooling takes place. The plant consisted of oil-filled electric radiators located at the appropriate zones. Each zone has a radiator with a heating capacity sufficient to bring the zone temperature to the heating set point. The dayroom was fitted with a 3000 W electric radiator, whereas the dining room was fitted with a 5000 W unit. Note that these values were arrived at by conducting an initial simulation without the plant. The heating output was split into 40% convective and 60% radiative.

It should be noted that the building, plant, and fluid flow arrangement was arrived at after a number of runs using different control schemes. The results for the predicted room temperature are shown in Fig. 4.

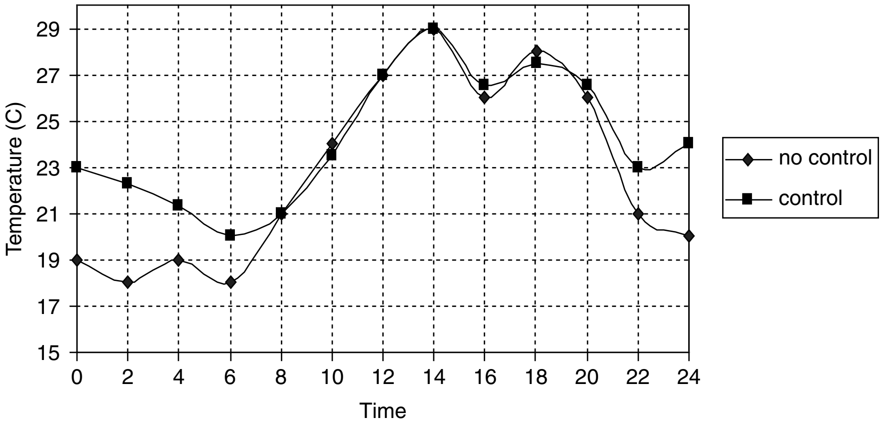
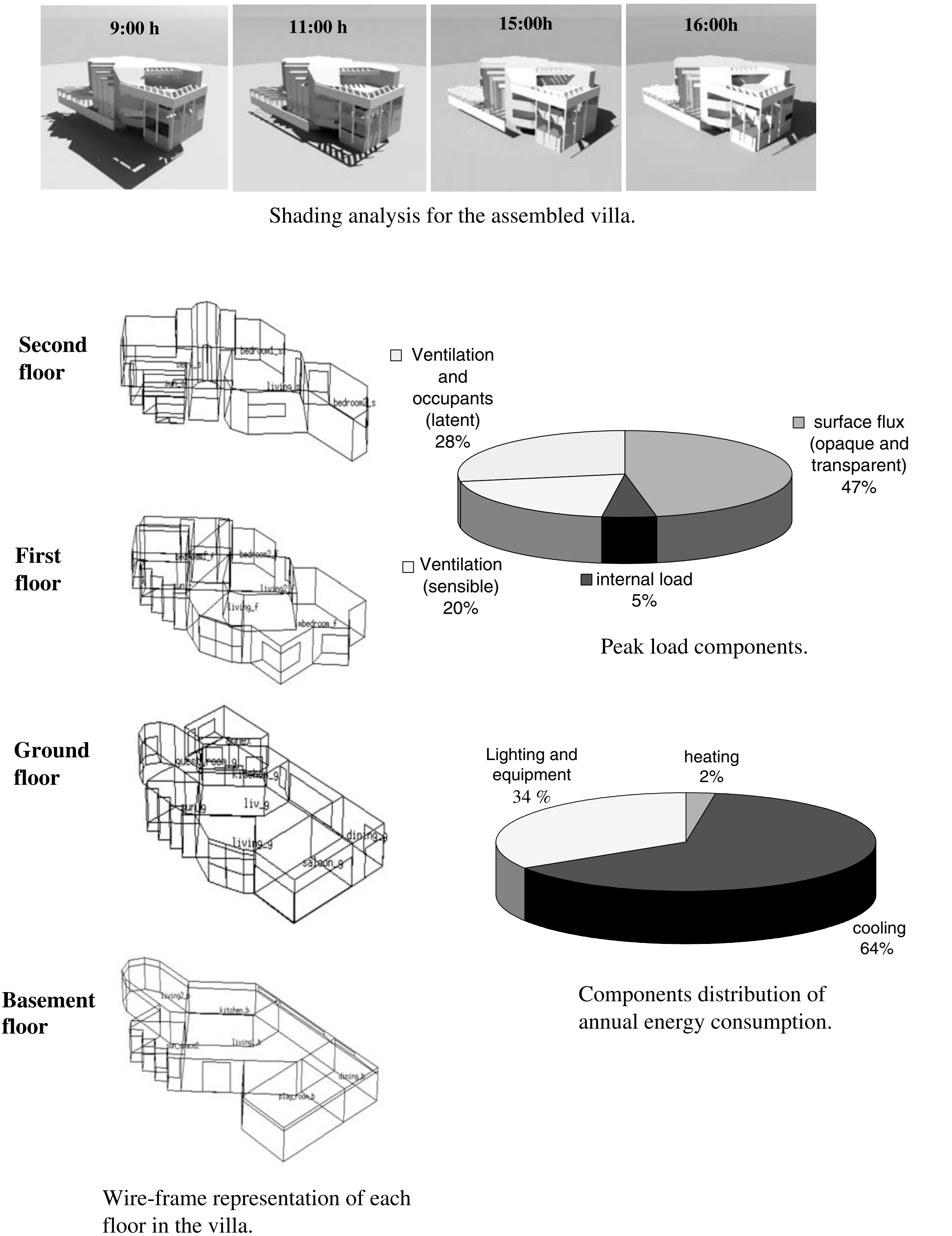


Fig. 4 Predicted dayroom temperature with plant and mass flow network with and without control on dayroom windows.



Shading analysis for the assembled villa.

Peak load components.

Components distribution of annual energy consumption.

Wire-frame representation of each floor in the villa.

Fig. 5 The different stages required, with detailed building energy analysis programs. Source: From Ref.1.

Fig. 4 shows that for the case with no control, the dayroom temperature is lower than the dining-room temperature when there is no solar radiation. The dayroom temperature increases during the day to a level slightly higher than the dining-room temperature. Heating takes place when the dayroom and dining-room temperatures fall below 21°C. A drop in the dayroom's upper-level temperature occurs in midmorning because that is when the extraction fan is activated by the thermostatic controller. With this building, plant, and flows arrangement, it was possible to decrease the dayroom temperature with respect to the base case. It also was possible to reduce the amount of heating required by the heating system if control also was imposed on the dayroom windows so that they remain closed if the dayroom temperature fell below a certain level. To achieve this, an additional flow component was considered in the simulation to represent flow control so that the window is closed if the dayroom temperature falls below 20°C. The predicted temperatures are also shown in Fig. 4 by the curve with control. The total heating required during the simulation period and in the case where the dayroom windows were left open was 102.2 kWhrs, compared with 39.5 kWhrs for the window-control case.

EXAMPLE 2

A private residential villa located in Kuwait City consists of large glazed areas and complex shading structures. For such villas, solar insulation is a concern because the weather in Kuwait City is typically hot, with a long summer extending over seven months when air conditioning is required. The number of cloudy days during the summer is almost negligible. Therefore, the electricity provider requested the investigation of the thermal performance of the villa when subjected to a number of energy conservation measures.[1] The plot has an area of 800 m^2, with the villa's constructional floor area being 1127 m^2, and is characterized by a sunspace. The external glazing area is 250 m^2 with a glazing to total external wall area ratio of 22%.

Fig. 5 shows that the villa has a complex geometrical shape with a complex shading structure. For this reason, a detailed shading analysis was carried out to account for the effect of the shading structures on solar insolation. In addition, annual simulations were conducted for the cases considered, which include:

- Base case: Use double tinted glass and no insulation on columns and beams.
- Case 1: Add thermal insulation to exposed floors.
- Case 2: Reduce window area by 20%.
- Case 3: Use double reflective glass.
- Case 4: Reduce air change by 50%.
- Case 5: Columns and beams fully insulated.
- Case 6: Use more energy efficient lighting.
- Case 7: Combine case-1, case-3 and case-5.
- Case 8: Combine case-4, and case-7.
- Case 9: Combine case-2, and case-8.
- Case 10: Combine case-2, and case-7.

For each case, the peak cooling demand and annual energy consumption were predicted by the BSS program. The peak cooling load was normalized over the living floor area so that it could be considered in updating the existing limit stipulated in the Kuwaiti code of practice for energy conservation in buildings.[11] It was possible from the simulation results to determine which conservation measure(s) to consider and what alterations were required so that the villa's thermal performance could be improved significantly over the base case design.

In addition, by including the cost factor, a cost/benefit analysis was performed, using the actual cost of implementing the energy conservation measure and using the simulation output to reflect the savings in terms of the initial cost of the air conditioning equipment and the cost of energy over the life cycle of the building. This gave a clear picture of the options that could be considered based on certain criteria (energy, cost, or both). A similar analysis was conducted on a number of other private villas, and the results of the normalized peak cooling load were used to update the limits for such buildings in the code of practice for energy conservation. This limit is of paramount importance because it is referred to by HVAC engineers to determine the maximum capacity of the air conditioning system; the limit can also be verified by the concerned authority for compliance with the code.

CONCLUSION

Building system simulation programs have evolved into more sophisticated tools over the past few decades. The developments made focused not only on the user interface level, but also on the mathematical approach adopted, leading to more accurate and realistic predictions of the thermal response of buildings. Building system simulation have wide applications for engineers, architects, and researchers. They are valuable tools that can be used to investigate the impact of numerous energy measures on the overall performance of buildings. State-of-the-art BSS tools allow for integrated building design, which consequently leads to designs that are optimized for thermal and visual comfort and for energy efficiency.

REFERENCES

1. Assem, E. Energy efficiency for fuelling the world. Proceedings of Energy Conservation in Buildings Workshop, Kuwait, December 13–17, Kwait Times: Kuwait, 2003, 305–328.

2. Ben-Nakhi, A. Development of an integrated dynamic thermal bridging assessment environment. Energy Buildings **2003**, *35* (4), 375–382.
3. Clarke, J.A. The energy kernel system. Energy Buildings **1988**, *10* (3), 259–266.
4. Crawley, D.B., et al. Energy plus: an update. Proceedings of SimBuild 2004 in Boulder, Colorado. International Building Performance Simulation Association: Boulder, Colorado, United States, August 4–6, 2004.
5. DOE-2 *DOE-2 Engineers Manual Version 2.1A*; National Technical Information Service, U.S. Department of Commerce: Springfield, Virginia, 1981.
6. ESP-r. The ESP-r System for Building Energy Simulation User Guide. Version 10 Series, ESRU Manual U02/1, University of Strathclyde: Glasgow, Scotland, October 2002.
7. Hand, J.; Assem, E. *Analysis of invernairn ward dayroom of erskine hospital*. ESRU Project Report, University of Strathclyde: Glasgow, Scotland, 1990.
8. Hensen, J.L.M., *On the Thermal Interaction of Building Structure and Heating and Ventilating Systems*. PhD thesis, Technische Universiteit: Eindhoven, Netherlands, 1991.
9. Clark, D.R., HVACSIM+ Program. Reference Manual. NBSIR 84-2996, National Institute of Standards and Testing, Gaithersberg, Maryland, 1985.
10. Kusuda, T. Thermal response factors for multilayer structures of various heat conduction systems. ASHRAE Trans. **1969**, *75* (1), 246–270.
11. Ministry of Electricity and Water. Code of practice for energy conservation in kuwaiti buildings. Report No. MEW R-6, State of Kuwait, 1983.
12. Negrao COR. *Conflation of Computational Fluid Dynamics and Building Thermal Simulation*. PhD thesis, University of Strathclyde: Glasgow, Scotland, 1995.
13. TRNSYS A transient simulation program. *Solar Energy Laboratory*, University of Wisconsin: Madison, Wisconsin, 1983.

Carbon Sequestration

Nathan E. Hultman
Science, Technology, and International Affairs, Georgetown University, Washington, D.C., U.S.A.

Bene—Clim

Abstract
Carbon dioxide (CO_2), a byproduct of hydrocarbon combustion and a natural emission from biomass burning, respiration, or decay, is a major greenhouse gas and contributor to anthropogenic climate change. Carbon sequestration describes the processes by which carbon can be either removed from the atmosphere (as CO_2) and stored, or separated from fuels or flue gases and stored. Carbon sequestration can thus be either technological (usually called carbon capture and storage) or biological (biological carbon sequestration). The viability of carbon sequestration depends on the cost of the process and the policy context that determines the value of sequestered carbon.

INTRODUCTION

The increasing likelihood of human-caused changes in climate could lead to undesirable impacts on ecosystems, economies, and human health and well-being. These potential impacts have prompted extensive assessment of options to reduce the magnitude and rate of future climate changes. Since climate changes are derived ultimately from increases in the concentrations of greenhouse gases (GHGs) in the atmosphere, such options must target either (a) reductions in the rate of inflow of GHGs to the atmosphere or (b) the removal of GHGs from the atmosphere once they have been emitted. Carbon sequestration refers to techniques from both categories that result in the storage of carbon that would otherwise be in the atmosphere as CO_2.

CO_2 is often targeted among the other GHGs because it constitutes the vast majority of GHG emissions by mass and accounts for three-fifths of the total anthropogenic contribution to climate change. Human emissions of CO_2 come primarily from fossil fuel combustion and cement production (80%), and land-use change (20%) that results in the loss of carbon from biomass or soil.

The rate of inflow of GHGs to the atmosphere can be reduced by a number of complementary options. For CO_2, mitigation options aim to displace carbon emissions by preventing the oxidation of biological or fossil carbon. These options include switching to lower-carbon fossil fuels, renewable energy, or nuclear power; using energy more efficiently; and reducing the rate of deforestation and land-use change. On the other hand, sequestration options that reduce emissions involve the capture and storage of carbon before it is released into the atmosphere.

Keywords: Carbon sequestration; Capture and storage; Carbon sinks; Geologic storage; Climate policy; Emissions trading; Carbon dioxide (CO_2).

CO_2 can also be removed directly from the atmosphere. While the idea of a large-scale, economically competitive method of technologically "scrubbing" CO_2 from the atmosphere is enticing, such technology currently does not exist. Policy has therefore focused on the biological process of carbon absorption through photosynthesis, either through expanding forested lands or, perhaps, enhancing photosynthesis in the oceans. This entry describes both the technological and biological approaches to carbon sequestration.

TECHNOLOGICAL SEQUESTRATION: CARBON CAPTURE AND STORAGE

The technological process of sequestering CO_2 requires two steps: first, the CO_2 must be separated from the industrial process that would otherwise emit it into the atmosphere; and second, the CO_2 must be stored in a reservoir that will contain it for a reasonable length of time. This process is therefore often referred to as carbon capture and storage (CCS) to distinguish it from the biological carbon sequestration that is described later.

Sources of Carbon

The best sites for CCS are defined by the efficiency of the capture technique, the cost of transport and sequestration, and the quantity of carbon available. The large capital requirements for CCS also dictate that large, fixed industrial sites provide the best opportunities. Therefore, although fossil-fueled transportation represents about 20% of current global CO_2 emissions, this sector presents no direct options for CCS at this time. The industrial sector, on the other hand, produces approximately 60% of current CO_2 emissions; most of these emissions come from large point sources which are ideal for CCS, such as power stations, oil refineries, petrochemical and gas reprocessing plants, and steel and cement works.[1]

Separation and Capture

Carbon capture requires an industrial source of CO_2; different industrial processes create streams with different CO_2 concentrations. The technologies applied to capture the CO_2 will therefore vary according to the specific capture process.[2–4] Capture techniques can target one of three sources:

- Post-combustion flue gases.
- Pre-combustion capture from gasification from power generation.
- Streams of highly pure CO_2 from various industrial processes.

Post-Combustion Capture

Conventional combustion of fossil fuels in air produces CO_2 streams with concentrations ranging from about 4 to 14% by volume. The low concentration of CO_2 in flue gas means that compressing and storing it would be uneconomical; therefore, the CO_2 needs to be concentrated before storage. Currently, the favored process for this task is chemical absorption, also known as chemical solvent scrubbing. Cooled and filtered flue gas is fed into an absorption vessel with a chemical solvent that absorbs the CO_2. The most common solvent for this process is monothanolamine (MEA). The CO_2-rich solvent is then passed to another reaction vessel called a stripper column. It is then heated with steam to reverse the process, thus regenerating the solvent and releasing a stream of CO_2 with a purity greater than 90%.

Scrubbing with MEA and other amine solvents imposes large costs in energy consumption in the regeneration process; it requires large amounts of solvents since they degrade rapidly; and it imposes high equipment costs since the solvents are corrosive in the presence of O_2. Thus, until solvents are improved in these areas, flue gas separation by this method will remain relatively costly: just the steam and electric load from a coal power plant can increase coal consumption by 40% per net kWh_e. Estimates of the financial and efficiency costs from current technology vary. Plant efficiency is estimated to drop from over 40% to a range between 24 and 37%.[2,5,6] For the least efficient systems, carbon would cost up to $70/t CO_2 and result in an 80% increase in the cost of electricity.[5] Other studies estimate an increase in the cost of electricity of 25%–75% for natural gas combined cycle and Integrated Gasification Combined Cycle (IGCC), and of 60%–115% for pulverized coal.[4] A small number of facilities currently practice flue gas separation with chemical absorption, using the captured CO_2 for urea production, foam blowing, carbonated beverages, and dry ice production. In addition, several developments may improve the efficiency of chemical absorption.

Several other processes have been proposed for flue-gas separation. Adsorption techniques use solids with high surface areas, such as activated carbon and zeolites, to capture CO_2. When the materials become saturated, they can be regenerated (releasing CO_2) by lowering pressure, raising temperature, or applying a low-voltage electric current. A membrane can be used to concentrate CO_2, but since a single pass through a membrane cannot achieve a great change in concentration, this process requires multiple passes or multiple membranes. An alternative use for membranes is to use them to increase the efficiency of the chemical absorption. In this case, a membrane separating the flue gas from the absorption solvent allows a greater surface area for the reaction, thus reducing the size and energy requirements of the absorption and stripper columns. *Cryogenic* techniques separate CO_2 from other gases by condensing or freezing it. This process requires significant energy inputs and the removal of water vapor before freezing.

One of the main limitations to flue-gas separation is the low pressure and concentration of CO_2 in the exhaust. An entirely different approach to post-combustion capture is to dramatically increase the concentration of CO_2 in the stream by burning the fuel in highly enriched oxygen rather than air. This process, called oxyfuel combustion, produces streams of CO_2 with a purity greater than 90%. The resulting flue gas will also contain some H_2O that can be condensed and removed, and the remaining high-purity CO_2 can be compressed for storage. Though significantly simpler on the exhaust side, this approach requires a high concentration of oxygen for the intake air. While this process alone may consume 15% of a plant's electric output, the separated N_2, Ar, and other trace gases also can be sold to offset some of the cost. Oxyfuel systems can be retrofitted onto existing boilers and furnaces.

Pre-Combustion Capture

Another approach involves removing the carbon from fossil fuels before combustion. First, the fuel is decomposed in the absence of oxygen to form a hydrogen-rich fuel called synthesis gas. Currently, this process of gasification is already in use in ammonia production and several commercial power plants fed by coal and petroleum byproducts; these plants can use lower-purity fuels and the energy costs of generating synthesis gas are offset by the higher combustion efficiencies of gas turbines; such plants are called IGCC plants. Natural gas can be transformed directly by reacting it with steam, producing H_2 and CO_2. While the principle of gasification is the same for all carbonaceous fuels, oil and coal require intermediate steps to purify the synthesis fuel and convert the byproduct CO into CO_2.

Gasification results in synthesis gas that contains 35%–60% CO_2 (by volume) at high pressure (over 20 bar). While current installations feed this resulting mixture into the gas turbines, the CO_2 can also be separated from the gas before combustion. The higher pressure and concentration give a CO_2 partial pressure of up to 50 times greater than in the post-combustion capture of flue gases, which enables another type of separation technique of physical solvent scrubbing. This technique is well known from ammonia production and involves the binding of CO_2 to solvents that release CO_2 in the stripper under lower pressure. Solvents in this category include cold methanol, polyethelene glycol, propylene carbonate, and sulpholane. The resulting separated CO_2 is, however, near atmospheric pressure and requires compression before storage (some CO_2 can be recovered at elevated pressures, which reduces the compression requirement). With current technologies, the total cost of capture for IGCC is estimated to be greater than $25 per ton of CO_2; plant efficiency is reduced from 43 to 37%, which raises the cost of electricity by over 25%.[5]

Pre-combustion capture techniques are noteworthy not only for their ability to remove CO_2 from fossil fuels for combustion in turbines, but also because the resulting synthesis gas is primarily H_2. They therefore could be an important element of a hydrogen-mediated energy system that favors the higher efficiency reactions of fuel cells over traditional combustion.[7]

Industrial CO_2 Capture

Many industrial processes release streams of CO_2 that are currently vented into the atmosphere. These streams, currently viewed as simple waste in an economically viable process, could therefore provide capture opportunities. Depending on the purity of the waste stream, these could be among the most economical options for CCS. In particular, natural gas processing, ethanol and hydrogen production, and cement manufacturing produce highly concentrated streams of CO_2. Not surprisingly, the first large-scale carbon sequestration program was run from a previously vented stream of CO_2 from the Sleipner gas-processing platform off the Norwegian coast.

Storage of Captured CO_2

Relatively small amounts of captured CO_2 might be re-used in other industrial processes such as beverage carbonation, mineral carbonates, or commodity materials such as ethanol or paraffins. Yet most captured CO_2 will not be re-used and must be stored in a reservoir. The two main routes for storing captured CO_2 are to inject it into geologic formations or into the ocean. However, all reservoirs have some rate of leakage and this rate is often not well known in advance. While the expected length of storage time is important (with targets usually in the 100–1000 year range), we must therefore also be reasonably confident that the reservoir will not leak more quickly than expected, and have appropriate measures to monitor the reservoir over time. Moreover, transporting CO_2 between the point of capture and the point of storage adds to the overall cost of CCS, so the selection of a storage site must account for this distance as well.

Geologic Sequestration

Geologic reservoirs—in the form of depleted oil and gas reservoirs, unmineable coal seams, and saline formations—comprise one of the primary sinks for captured CO_2. Estimates of total storage capacity in geologic reservoirs could be up to 500% of total emissions to 2050 (Fig. 1).

Captured CO_2 can be injected into depleted oil and gas reservoirs, or can be used as a means to enhance oil recovery from reservoirs nearing depletion. Because they held their deposits for millions of years before extraction, these reservoirs are expected to provide reliable storage for CO_2. Storage in depleted reservoirs has been practiced for years for a mixture of petroleum mining waste gases called "acid gas."

A petroleum reservoir is never emptied of all its oil; rather, extracting additional oil just becomes too costly to justify at market rates. An economically attractive possibility is therefore using captured CO_2 to simultaneously increase the yield from a reservoir as it is pumped into the reservoir for storage. This process is called enhanced oil recovery. Standard oil recovery yields only about 30%–40% of the original petroleum stock. Drilling companies have years of experience with using compressed CO_2, a hydrocarbon solvent, to obtain an additional 10%–15% of the petroleum stock. Thus, captured CO_2 can be used to provide a direct economic

Reservoir Type	Storage Capacity	
	billion tonnes CO_2	% of E
Coal basins	170	8%
Depleted oil reservoirs	120	6%
Gas basins	700	34%
Saline formations		
Terrestrial	5,600	276%
Off-shore	3,900	192%
Total	10,490	517%

Fig. 1 CO_2 Reservoirs. Carbon dioxide storage capacity estimates. E is defined as the total global CO_2 emissions from the years 2000–2050 in IPCC's business-as-usual scenario IS92A. Capacity estimates such as these are rough guidelines only and actual utilization will depend on carbon economics. Source: From Ref. 8.

benefit along with its placement in a reservoir. This benefit can be used to offset capture costs.

Coal deposits that are not economically viable because of their geologic characteristics provide another storage option. CO_2 pumped into these unmineable coal seams will adsorb onto the coal surface. Moreover, since the coal surface prefers to adsorb CO_2 to methane, injecting CO_2 into coal seams will liberate any coal bed methane (CBM) that can then be extracted and sold. This enhanced methane recovery is currently used in U.S. methane production, accounting for about 8% in 2002. Such recovery can be used to offset capture costs. One potential problem with this method is that the coal, as it adsorbs CO_2, tends to swell slightly. This swelling closes pore spaces and thus decreases rock permeability, which restricts both the reservoir for incoming CO_2 and the ability to extract additional CBM.

Saline formations are layers of porous sedimentary rock (e.g., sandstone) saturated with saltwater, and exist both under land and under the ocean. These layers offer potentially large storage capacity representing several hundred years' worth of CO_2 storage. However, experience with such formations is much more limited and thus the uncertainty about their long-term viability remains high. Moreover, unlike EOR or CBM recovery with CO_2, injecting CO_2 into saline formations produces no other commodity or benefit that can offset the cost. On the other hand, their high capacity and relative ubiquity makes them attractive options in some cases. Statoil's Sleipner project, for example, uses a saline aquifer for storage.

Research and experimentation with saline formations is still in early stages. To achieve the largest storage capacities, CO_2 must be injected below 800 m depth, where it will remain in a liquid or supercritical dense phase (supercritical point at 31°C, 71 bar). At these conditions, CO_2 will be buoyant (a density of approximately 600–800 kg/m^3) and will tend to move upward. The saline formations must therefore either be capped by a less porous layer or geologic trap to prevent leakage of the CO_2 and eventual decompression.[9] Over time, the injected CO_2 will dissolve into the brine and this mixture will tend to sink within the aquifer. Also, some saline formations exist in rock that contains Ca-, Mg-, and Fe-containing silicates that can form solid carbonates with the injected CO_2. The resulting storage as rock is highly reliable, though it may also hinder further injection by closing pore spaces. Legal questions may arise when saline formations, which are often geographically extensive, cross national boundaries or onto marine commons.

Ocean Direct Injection

As an alternative to geologic storage, captured CO_2 could be injected directly into the ocean at either intermediate or deep levels. The oceans have a very large potential for storing CO_2, equivalent to that of saline aquifers ($\sim 10^3$ Gt). While the ocean's surface is close to equilibrium with atmospheric carbon dioxide concentrations, the deep ocean is not because the turnover time of the oceans is much slower ($\sim$5000 years) than the observed increases in atmospheric CO_2. Since the ocean will eventually absorb much of the atmospheric perturbation, injecting captured CO_2 into the oceans can therefore be seen as simply bypassing the atmospheric step and avoiding the associated climate consequences. Yet little is known about the process or effects—either ecological or geophysical—of introducing large quantities of CO_2 into oceanic water.

At intermediate depths (between 500 and 3000 m), CO_2 exists as a slightly buoyant liquid. At these depths, a stream of CO_2 could be injected via a pipe affixed either to ship or shore. The CO_2 would form a droplet plume, and these droplets would slowly dissolve into the seawater, disappearing completely before reaching the surface. Depressed pH values are expected to exist for tens of km downcurrent of the injection site, though changing the rate of injection can moderate the degree of perturbation. In addition, pulverized limestone could be added to the injected CO_2 to buffer the acidity.

Below 3000 m, CO_2 becomes denser than seawater and would descend to the seafloor and pool there. Unlike intermediate injection, therefore, this method does not lead to immediate CO_2 dissolution in oceanic water; rather, the CO_2 is expected to dissolve into the ocean at a rate of about 0.1 m/y. Deep injection thus minimizes the rate of leakage to the surface, but could still have severe impacts on bottom-dwelling sea life.

The primary obstacles to oceanic sequestration are not technical but relate rather to this question of environmental impacts.[10] Oceanic carbon storage might affect marine ecosystems through the direct effects of a lower environmental pH; dissolution of carbonates on fauna with calcareous structures and microflora in calcareous sediments; impurities such as sulfur oxides, nitrogen oxides, and metals in the captured CO_2; smothering effects (deep injection only); and changes in speciation of metals and ammonia due to changes in pH. Few of these possibilities have been studied in sufficient detail to allow an informed risk assessment. In addition, the legality of dumping large quantities of CO_2 into the open ocean remains murky.

Overall Costs of CCS

The costs of CCS can be measured either as a cost per tonne of CO_2, or, for power generation, a change in the cost of electricity (Fig. 2). The total cost depends on the cost of capture, transport, and storage. Capture cost is mainly a function of parasitic energy losses and the capital cost of equipment. Transport cost depends on distance and terrain. Storage costs vary depending on the reservoir but are currently a few dollars per tonne of CO_2. The variety of

Fossil plant type	Cost of CCS
	¢ per kWh
Natural gas combined cycle	1–2
Pulverized coal	2–3
Coal IGCC	2–4

Fig. 2 Additional costs to power generation from CCS. Approximate capture and storage costs for different approaches to power plant sequestration.
Source: From Refs. 4,5,11,12.

approaches to CCS and the early stages of development make precise estimates of cost difficult, but current technology spans about \$25–\$85/t CO_2.

BIOLOGICAL SEQUESTRATION: ENHANCING NATURAL CARBON SINKS

The previous sections have described processes by which CO_2 could be technologically captured and then stored. Photosynthesis provides an alternate route to capture and store carbon. Enhancing this biological process is therefore an alternative method of achieving lower atmospheric CO_2 concentrations by absorbing it directly from the air.

Terrestrial Carbon Sinks

Carbon sequestration in terrestrial ecosystems involves enhancing the natural sinks for carbon fixed in photosynthesis. This occurs by expanding the extent of ecosystems with a higher steady-state density of carbon per unit of land area. For example, because mature forest ecosystems contain more carbon per hectare than grasslands, expanding forested areas will result in higher terrestrial carbon storage. Another approach is to encourage the additional storage of carbon in agricultural soils. The essential element in any successful sink enhancement program is to ensure that the fixed carbon remains in pools with long lives.

Afforestation involves planting trees on unforested or deforested land.[13,14] The most likely regions for forest carbon sequestration are Central and South America and Southeast Asia because of relatively high forest growth rates, available land, and inexpensive labor. However, the translation of forestry activities into a policy framework is complex. Monitoring the carbon changes in a forest is difficult over large areas, as it requires not only a survey of the canopy and understory, but also an estimate of the below-ground biomass and soil carbon. Some groups have voiced concern over the potential for disruption of social structures in targeted regions.

Soil carbon sequestration involves increasing soil carbon stocks through changes in agriculture, forestry, and other land use practices. These practices include mulch farming, conservation tillage, agroforestry and diverse cropping, cover crops, and nutrient management that integrates manure, compost, and improved grazing. Such practices, which offer the lowest-cost carbon sequestration, can have other positive effects such as soil and water conservation, improved soil structure, and enhanced soil fauna diversity. Rates of soil carbon sequestration depend on the soil type and local climate, and can be up to 1000 kg of carbon per hectare per year. Management practices can enhance sequestration for 20–50 years, and sequestration rates taper off toward maturity as the soil carbon pool becomes saturated. Widespread application of recommended management practices could offset 0.4 to 1.2 GtC/y, or 5%–15% of current global emissions.[15]

If sinks projects are to receive carbon credits under emissions trading schemes like that in the Kyoto Protocol, they must demonstrate that the project sequestered more carbon than a hypothetical baseline or business-as-usual case. They must also ensure that the carbon will remain in place for a reasonable length of time, and guard against simply displacing the baseline activity to a new location.

Ocean Fertilization

Vast regions of the open ocean have very little photosynthetic activity, though sunlight and major nutrients are abundant. In these regions, phytoplankton are often deprived of trace nutrients such as iron. Seeding the ocean surface with iron, therefore, might produce large phytoplankton blooms that absorb CO_2. As the plankton die, they will slowly sink to the bottom of the ocean, acting to transport the fixed carbon to a permanent burial in the seafloor. While some experimental evidence indicates this process may work on a limited scale, little is known about the ecosystem effects and potential size of the reservoir.[16]

PROSPECTS FOR CARBON SEQUESTRATION

Carbon sequestration techniques—both technological and biological—are elements of a portfolio of options for addressing climate change. Current approaches hold some promise for tapping into the geologic, biologic, and oceanic potential for storing carbon. The costs of some approaches, especially the improved management of agricultural and forest lands, are moderate (Fig. 3).

Sequestration Technique	Cost
	\$ per T CO_2
Carbon capture & storage	26–84
Tree planting & agroforestry	10–210
Soil carbon sequestration	6–24

Fig. 3 Costs of carbon sequestration. Estimates for sequestration costs vary widely. Future costs will depend on rates of technological change.
Source: From Refs. 4,5,7,11–14.

Bene—Clim

Yet these opportunities are not infinite and additional options will be necessary to address rising global emissions. Thus, the higher costs of current technological approaches are likely to drop with increasing deployment and changing market rates for carbon.

Possible developments include advanced CO_2 capture techniques focusing on membranes, ionic (organic salt) liquids, and microporous metal organic frameworks. Several alternative, but still experimental, sequestration approaches have also been suggested. Mineralization could convert CO_2 to stable minerals. This approach seeks, therefore, to hasten what in nature is a slow but exothermic weathering process that operates on common minerals like olivine, forsterite, or serpentines (e.g., through selected sonic frequencies). It is possible that CO_2 could be injected in sub-seafloor carbonates. Chemical looping describes a method for combusting fuels with oxygen delivered by a redox agent instead of by air or purified oxygen; it promises high efficiencies of energy conversion and a highly enriched CO_2 exhaust stream. Research also continues on microbial CO_2 conversion in which strains of microbes might be created to metabolize CO_2 to produce saleable commodities (succinic, malic, and fumeric acids). In addition, the nascent science of monitoring and verifying the storage of CO_2 will be an important element toward improving technical performance and public acceptance of sequestration techniques.

ACKNOWLEDGMENTS

The author thanks the anonymous reviewer for helpful comments on an earlier draft.

REFERENCES

1. Gale, J., Overview of CO_2 Emission Sources, Potential, Transport, and Geographical Distribution of Storage Possibilities. In *Proceedings of IPCC workshop on carbon dioxide capture and storage, Regina, Canada*, Nov 18–21, 2002; 15–29. http://arch.rivm.nl/env/int/ipcc/pages_media/ccs2002.html (accessed April 2005). A revised and updated version of the papers presented at the Regina workshop is now available. See, Metz, B.; Davidson, O.; de Coninck, H.; Loos, M.; Meyer, L., Special Report on Carbon Dioxide Capture and Storage. Intergovernmental Panel on Climate Change: Geneva, http://www.ipcc.ch. (accessed May 2006), 2006.
2. Thambimuthu, K.; Davison, J.; Gupta, M., CO_2 Capture and Reuse, In *Proceedings of IPCC workshop on carbon dioxide capture and storage*, Regina, Canada, Nov 18–21, 2002; 31–52, http://arch.rivm.nl/env/int/ipcc/pages_media/ccs2002.html (accessed April 2005).
3. Gottlicher, G. *The Energetics of Carbon Dioxide Capture in Power Plants*; U.S. Department of Energy Office of Fossil Energy: Washington, DC, 2004; 193.
4. Herzog, H.; Golomb, D. Carbon capture and storage from fossil fuel use. In *Encyclopedia of Energy*; Cleveland, C.J., Ed.; Elsevier Science: New York, 2004; 277–287.
5. National Energy Technology Laboratory *Carbon Sequestration: Technology Roadmap and Program Plan*; U.S. Department of Energy: Washington, DC, 2004.
6. Gibbins, J.R.; Crane, R.I.; Lambropoulos, D.; Booth, C.; Roberts, C.A.; Lord, M., Maximizing the Effectiveness of Post Combustion CO_2 Capture Systems. In *Proceedings of the 7th International Conference on Greenhouse Gas Abatement*, Vancouver, Canada, Sept 5–9, 2004; Document E2-2. http://www.ghgt7.ca (accessed June 2004).
7. Williams, R.H., Decarbonized Fossil Energy Carriers and Their Energy Technology Competitors. In *Proceedings of IPCC workshop on carbon dioxide capture and storage*, Regina, Canada, November, 2002; 119–135. http://arch.rivm.nl/env/int/ipcc/pages_media/ccs2002.html (accessed April 2005).
8. Dooley, J.J.; Friedman, S.J. *A Regionally Disaggregated Global Accounting of CO_2 Storage Capacity: Data and Assumptions*; Battelle National Laboratory: Washington DC, 2004; 15.
9. Kårsted, O., Geological Storage, Including Costs and Risks, in Saline Aquifers. In *Proceedings of IPCC workshop on carbon dioxide capture and storage*, Regina, Canada, Nov 18–21, 2002; 53–60. http://arch.rivm.nl/env/int/ipcc/pages_media/ccs2002.html (accessed April 2005).
10. Johnston, P.; Santillo, D., Carbon Capture and Sequestration: Potential Environmental Impacts, In *Proceedings of IPCC workshop on carbon dioxide capture and storage*, Regina, Canada, Nov 18–21, 2002; 95–110. http://arch.rivm.nl/env/int/ipcc/pages_media/ccs2002.html (accessed April 2005).
11. Dooley, J.J.; Edmonds, J.A.; Dahowski, R.T.; Wise, M.A., Modeling Carbon Capture and Storage Technologies in Energy and Economic Models. In *Proceedings of IPCC workshop on carbon dioxide capture and storage*, Regina, Canada, Nov 18–21, 2002; 161–172. http://arch.rivm.nl/env/int/ipcc/pages_media/ccs2002.html (accessed April 2005).
12. Freund, P.; Davison, J., General Overview of Costs. In *Proceedings of IPCC workshop on carbon dioxide capture and storage*, Regina, Canada, Nov 18–21, 2002; 79–93. http://arch.rivm.nl/env/int/ipcc/pages_media/ccs2002.html (accessed April 2005).
13. Van Kooten, G.C.; Eagle, A.J.; Manley, J.; Smolak, T. How costly are carbon offsets? A meta-analysis of carbon forest sinks. Environ. Sci. Policy **2004**, *7*, 239–251.
14. Richards, K.R.; Stokes, C. A review of forest carbon sequestration cost studies: a dozen years of research. Climatic Change **2004**, *68*, 1–48.
15. Lal, R. Soil carbon sequestration impacts on global climate change and food security. Science **2004**, *304*, 1623–1627.
16. Buesseler, K.O.; Andrews, J.E.; Pike, S.M.; Charette, M.A. The effects of iron fertilization on carbon sequestration in the southern ocean. Science **2004**, *304*, 414–417.

Career Advancement and Assessment in Energy Engineering

Albert Thumann
Association of Energy Engineers, Atlanta, Georgia, U.S.A.

Abstract
The Association of Energy Engineers (AEE) has helped define the profession of energy engineering through continuing education programs and journals. Association of Energy Engineers provides many networking opportunities through local chapters. This entry details continuing education programs available through AEE, the salary structure of energy professionals, and achieving excellence through certifications.

INTRODUCTION

The profession of energy engineering has gained new importance for several important factors. First, global warming is now considered a reality, and there is a need to reduce greenhouse gases by applying energy-efficient technologies. In addition, surging energy prices have caused companies to evaluate how they use energy and to reduce operating costs. A third factor is the reliability of the electric power grid. In the summer of 2006, blackouts occurred in Queens, New York and other parts of the country. Energy engineering professionals play a key role in reducing the need for power generation and distribution lines. The Association of Energy Engineers (AEE) is dedicated to improving the practice of energy engineering through certification programs, continuing education programs, and networking opportunities through 44 chapters throughout the world.

The profession of energy engineering is relatively new. The oil embargo of the early 1970s created a demand for engineers who can apply the latest energy efficiency technologies to reduce demand in buildings and industrial processes. Most colleges do not offer degrees in energy engineering. An individual pursuing this profession usually attends courses presented by universities and associations, and receives on-the-job training. Some universities—such as Texas A&M, Georgia Tech, and the University of Wisconsin—offer graduate courses in energy engineering and management.

THE ASSOCIATION OF ENERGY ENGINEERS

The AEE was founded in 1977 as a 501 c(6) not-for-profit professional society. The purpose of the AEE is to promote the scientific and educational aspects of those engaged in the energy industry. In the 1970s, the profession of energy engineering and energy management was new. The AEE defined the important functions energy engineers and managers perform and played a key role in the professions' development.

One of the AEE's first tasks was to create an authoritative journal that would guide energy engineers in applying new energy-efficient technologies and applications. The *Energy Engineering* journal was born out of this need and currently is edited by noted authority Dr. Wayne Turner.

The AEE recognized that energy engineers need both technical and management skills. Energy engineers need a broad understanding of fuels procurement, commodity, and risk management, as well as organizational and motivational skills. The *Strategic Planning for Energy and the Environment* journal, also edited by Dr. Wayne Turner, was developed by the AEE to meet this need.

To help energy engineers meet the challenges of power reliability and the development of new energy supplies, the AEE launched the *Cogeneration and Distributed Generation* journal, currently edited by Dr. Steven Parker.

Today, the AEE's network includes 8500 members in 77 countries, with chapters in 69 cities.

The AEE presents numerous training and certification programs to help energy engineers reach their potential.

For complete details on programs offered by the AEE, the reader is referred to www.aeecenter.org.

SALARY STRUCTURE OF ENERGY ENGINEERS

The AEE conducted a salary survey of its 8500 members in 2005. The results of the salary survey follow.

AEE INCOME AND SALARY SURVEY

Only those who were employed full time as of January 1, 2005 were surveyed.

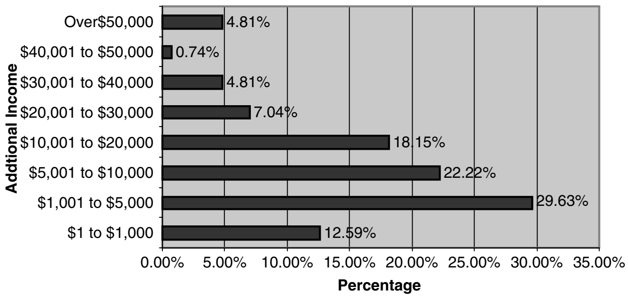
Keywords: Profession; Energy engineering; Association of Energy Engineers (AEE); Salary structure of energy engineers; Continuing education programs; Continuing Education Units (CEUs); Scholarship; Professional certification programs.

1. Please input your base salary (to the nearest $10,000) as of January 1, 2004, to January 1, 2005 (exclude bonus, overtime, fees, and income from secondary employment).

Base Salary	Total	Percentage
$20,000 to $30,000	3	0.65%
$30,000 to $40,000	5	1.08%
$40,000 to $50,000	17	3.66%
$50,000 to $60,000	47	10.13%
$60,000 to $70,000	82	17.67%
$70,000 to $80,000	99	21.34%
$80,000 to $90,000	78	16.81%
$90,000 to $100,000	56	12.07%
$100,000 to $110,000	36	7.76%
$110,000 to $120,000	13	2.80%
$120,000 to $130,000	9	1.94%
$130,000 to $140,000	5	1.08%
$140,000 to $150,000	5	1.08%
Over $150,000	9	1.94%

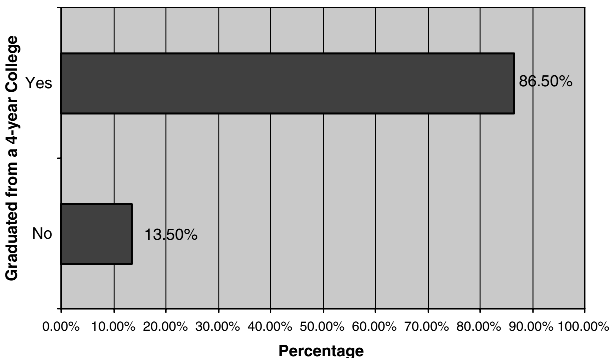
Average Annual Salary: $85,625.00.

2. Please input your additional income (to the nearest $1000) from primary job, such as bonus, overtime, and fees as of January 1, 2004 to January 1, 2005.

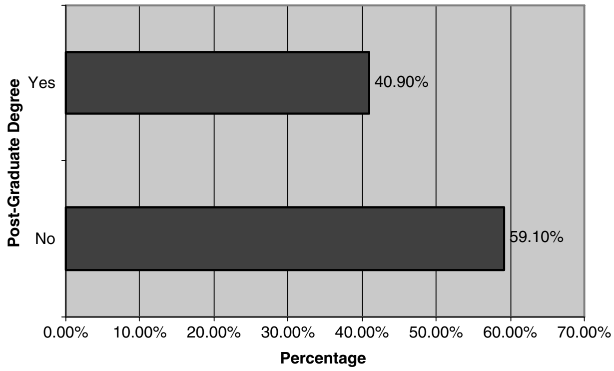


Average Annual Bonus Amount: $14,274.07.

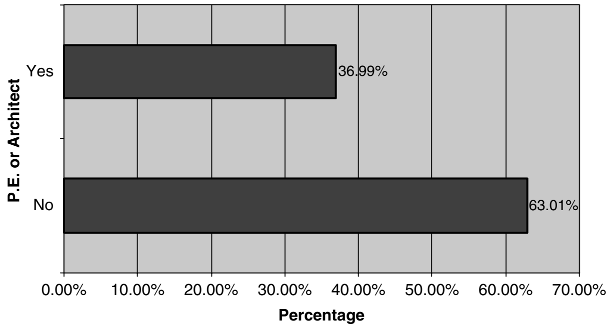
3. Are you a graduate from a 4-year accredited college?



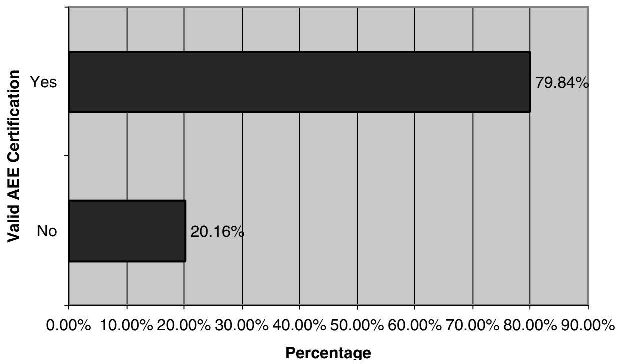
4. Do you have a post-graduate degree from an accredited college?



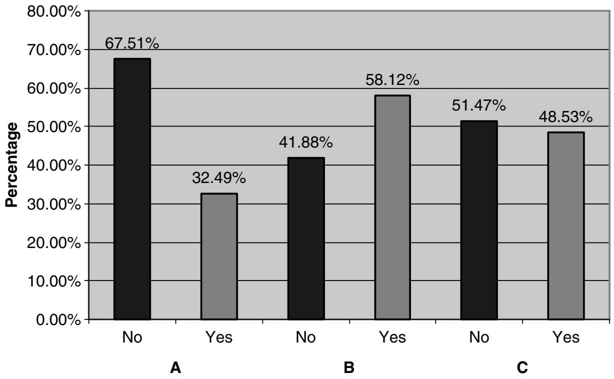
5. Are you a registered Professional Engineer or Architect?



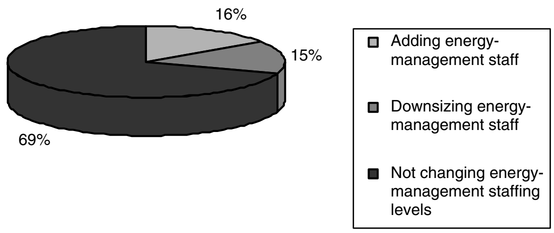
6. Do you hold a valid certification from AEE, such as a Certified Energy Manager (CEM), Energy Manager In Training (EMIT), Certified Lighting Efficiency Professional (CLEP), Certified Power Quality (CPQ), Certified Building Commissioning Professional (CBCP), Distributed Generation Certified Professional (DGCP), Certified Measurement & Verification Professional (CMVP), Certified Demand Side Manager (CDSM), Certified Cogeneration Professional (CCP), Certified Energy Procurement (CEP), Certified Indoor Air Quality Professional (CIAQP), Certified Indoor Air Quality Technician (CIAQT), Certified Testing, Adjusting and Balancing (CTAB) or Certified GeoExchange Designer (CGD)?



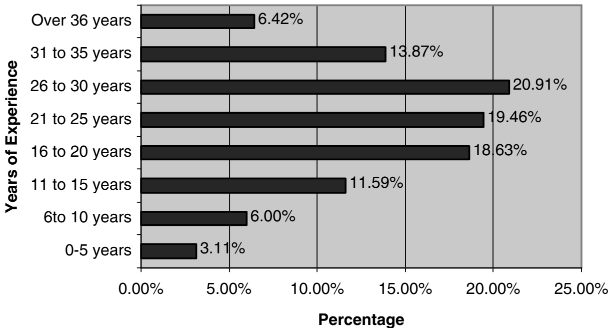
7. Because you have assumed energy management responsibilities at your company, are you:
   A Receiving significantly higher compensation than before?
   B Receiving higher visibility?
   C In a better position for advancement?



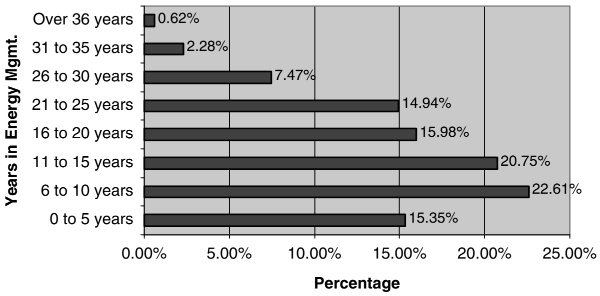
8. Is your company currently:



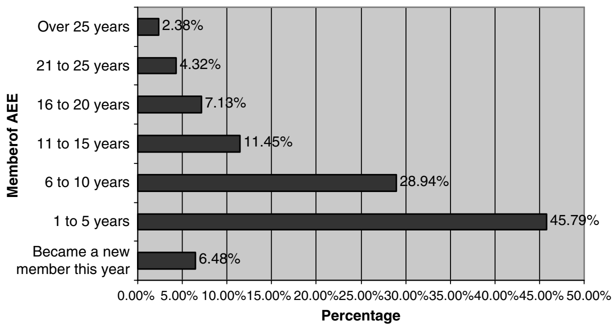
9. How many years of experience do you have?



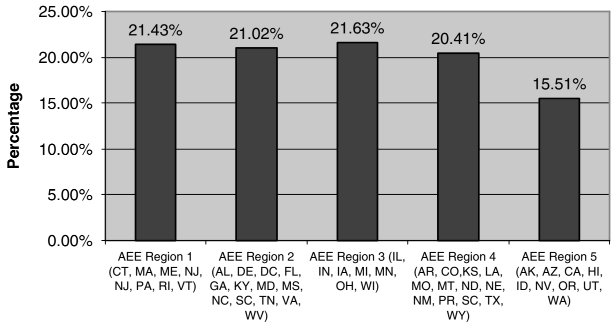
10. How many of those years have you been involved in energy management?



11. How many years have you been a member of the Association of Energy Engineers?



12. Please identify the location where you are employed.



PROFESSIONAL CERTIFICATION PROGRAMS

Most professionals practicing in the energy engineering industry do not have a degree in this field. Specialty certification programs developed by the AEE and other professional organizations offer the following benefits:

- Certification is part of total quality management. When an individual becomes certified in a designated field, his or her professional achievement is recognized in the eyes of colleagues, government agencies, present and prospective employers, and clients.
- Certification establishes a standard of professional competence that is recognized throughout the industry.
- Certification fosters development of the individual's capabilities through encouragement of long-term career goals.
- Certification promotes quality through continuing education to ensure a high level of competence within constantly changing fields.

CERTIFICATION PROGRAMS OFFERED BY THE AEE

The Certified Energy Manager (CEM) program is broad based and emphasizes the technical aspects of what energy managers in government and the private sector need to know. The Certified Energy Manager represents a "who's who" of energy management. Since 1982, more than 6400 individuals have gained the status of CEM.

The Certified Business Energy Professional (BEP) program awards special recognition to those business/marketing and energy professionals who have demonstrated a high level of competence and ethical fitness for business/marketing and energy management-related disciplines, as well as laws governing and affecting energy professionals.

The Certified Lighting Efficiency Professional (CLEP) program was developed to identify specialists in lighting efficiency. This program is also recognized as meeting the certification requirements of the U.S. Environmental Protection Agency's Green Lights Lighting Management Company Ally program.

The Certified Indoor Air Quality Professional (CIAQP) program is designed to meet the growing needs of businesses to identify qualified indoor-air-quality practitioners who are equipped to solve problems created by "sick buildings," as well as facility managers who are responsible for operating healthy buildings while maintaining comfort and reducing energy costs.

The Certified Power Quality Professional (CPQ) program demonstrates the interactions between the power source and sensitive loads in the field of power quality and reliability.

The Certified Geoexchange Designer (CGD) program is designed to recognize professionals who have demonstrated high levels of experience, competence, proficiency, and ethical fitness in applying the principles and practices of geothermal heat pump design and related disciplines. The CGD certification is granted by the AEE and sponsored by the Geothermal Heat Pump Consortium (GHPC). Associated training programs are presented by the International Ground Source Heat Pump Association (IGSHPA).

The Certified Energy Procurement Professional (CEP) program covers the acquisition of both electricity and natural gas from the purchasing/procurement and selling/marketing perspectives.

The Distributed Generation Certified Professional (DGCP) program identifies individuals who have demonstrated high levels of experience, competence, proficiency, and ethical fitness, bringing to their professional activities the full scope of knowledge essential to the effective development and management of today's distributed generation projects.

The Certified Measurement and Verification Professional (CMVP) program was established by the AEE in cooperation with the International Performance Measurement and Verification Protocol (IPMVP), with the dual purpose of recognizing the most qualified professionals in this growing area of the energy industry, as well as raising overall professional standards within the measurement and verification field.

The Certified Building Commissioning Professional (CBCP) program was developed with the dual purpose of recognizing the most highly qualified professionals in this rapidly expanding area within the industry and raising overall professional standards in the building-commissioning field.

The Certified Hotel Environmental Manager (CHEM) program is designed for the multiple purposes of raising the professional standards of those engaged in hotel environmental management, identifying individuals with acceptable knowledge of the principles and practices of hotel environmental management, and awarding those individuals special recognition for demonstrating a high level of competence and ethical fitness in hotel environmental management. Sponsored by the U.S. Agency for International Development (USAID), the CHEM certification is granted by the AEE. The associated training program is presented by PA Consulting Group.

The Emissions Trading Certified Professional (ETC) program is designed to award special recognition to professionals who have demonstrated a high level of knowledge, experience, competence, and ethical fitness covering the full spectrum of activities related to the trading of emissions allowances and evaluation of emissions credits. By obtaining the ETC credential, the individual establishes his or her status as a qualified expert in this growing area of specialized expertise.

THE AEE CERTIFICATION PROCESS

Each program requires the individual to complete an application, which includes requests for the following information:

- Demonstrated experience in the field. Each program specifies the minimum experience levels required. Employee or client verification is also required.
- Documented education or professional registration.
- Professional references to verify experience and qualifications.

In addition, the individual must complete a four-hour open-book examination. The exam consists of multiple-choice and true/false questions. The actual test questions are framed to ascertain both specific knowledge and practical experience. Sample questions and study guides are available from the AEE. The completed application and test score weigh equally in determining whether the individual meets certification requirements;

both are reviewed by the governing board of the specific program.

CONTINUING EDUCATION PROGRAMS OFFERED BY AREA

The AEE offers a wide range of training options. Each training option offers Continuing Education Units (CEUs), which are important for documenting courses successfully completed (one CEU equals ten professional development hours [pdh]). In addition, a certificate of participation is awarded.

In 2005, 27 states required CEUs as a prerequisite for renewal of professional engineering licenses.

Programs offered by the AEE include:

- *Live seminars*. The AEE presents a wide range of courses in cities across the nation. Several seminars are designed to prepare students for the professional certification examinations. Live programs offer an optimum learning environment with ample time to interact with the instructor, as well as colleagues in attendance.
- *In-house seminars*. Most of the live seminars, including professional certification training programs, can also be presented at company facilities around the world.
- *Online real-time training (synchronous)*. Students can participate in a real-time seminar at the office or at home with access to the Internet and a telephone. Students communicate with the instructor through a scheduled conference call and view the instructor's Microsoft PowerPoint presentation via the Internet.
- *Self-paced online training (asynchronous)*. Each student receives a workbook containing the training materials and examination questions and completes the training at his or her own pace. The students who pass the online examination receive a certificate of course completion and are awarded CEUs. Students can interact with the instructor and fellow students during regularly scheduled chat sessions.
- *24/7 online training (asynchronous)*. The course material is accessed online 24 hr a day for up to 30 days. The students who pass the online examination are able to print a certificate of course completion and are awarded CEUs.
- *Conferences and expositions*. The AEE offers three conferences and expositions each year. The purpose is to present the latest technologies and applications by leading experts in the field. The flagship event presented by the AEE is the World Energy Engineering Congress (WEEC).

STUDENT PROGRAMS FOR THE ENERGY ENGINEERING PROFESSION THROUGH THE AEE

The future of the energy engineering profession is in developing new talent through student programs. The AEE encourages student participation as follows:

- *Student membership*. Student members receive all publications and can participate at a reduced rate of $15 annually. Dues are subsidized by the AEE.
- *Student chapters*. Networking with fellow classmates and seasoned professions is accomplished through student chapters across the nation.
- *Scholarships*. To help students further their educations in the field of energy engineering, scholarships are offered. The Foundation of the AEE is a 501 (c)3 not-for-profit corporation. Since its inception, the AEE has awarded $480,000 in scholarships.

VISIONS FOR THE FUTURE

The energy engineering profession continues to grow. The need for energy efficiency has never been greater, due to the following circumstances.

ENERGY SECURITY

- Since 1985, imports of refined petroleum products have increased by 34%. Today, the total import of oil is 56.8%. The volatility in the Middle East and other foreign sources of supply has led to disruptions in the availability of oil and higher prices. By mid-2005, oil prices spiraled to more than $60 per barrel.
- Since 1970, U.S. production of crude oil has declined from 9.6 to 5.8 million barrels per day. While, consumption has increased from 14.7 to 20 million barrels per day. The Organization of Petroleum Exporting Countries (OPEC) has tightened supplies in the past, causing gasoline and oil prices to spiral. The "war on terrorism" will cause further instability in the Middle East and prioritize this nation's need to be energy secure.
- The power blackout of 2003 affecting New York, Detroit, and other major cities indicates that the present transmission grid is congested, outdated, and in dire need of overhaul. According to the U.S. Energy Information Administration, electricity use will increase 22% by 2010, placing further demands on an obsolete transmission grid.
- The long-term price of natural gas in the United States has more than doubled over the past six years. U.S.

natural gas production history shows that new wells are being depleted more quickly all the time; the current decline rate is 28% per year. Although this is partially due to growing demand, it is also due to the fact that the large fields of natural gas are all aging and in terminal decline. Newer natural gas fields tend to be smaller and are produced (and depleted) quickly in the effort to maintain overall production levels. Production from wells drilled in 2003 has been declining at a rate of 23% per year.

ENERGY-EFFICIENT ECONOMY

Energy efficiency improvements have had a major impact on companies' profitability, decreasing energy usage, and reducing greenhouse gasses. Over the past 25 years, per capita use of energy has declined 0.8%. A continuing drive for energy efficiency can help keep prices down and buy the nation time to address critical supply problems.

Energy efficiency programs have saved consumers more than $25 billion a year while improving the quality of life. Energy efficiency saved 70 quadrillion Btu from 1972 to 1999.

Energy efficiency technologies can also give the nation time to rebuild and modernize the electric transmission infrastructure.

UPGRADING TRANSMISSION AND DISTRIBUTION SYSTEMS

The U.S. electric generation and transmission system is 70 years old and is based on technologies from the 1950s. The electric transmission grid consists of approximately 160,000 mi of high-voltage transmission lines and is in dire need of replacement and expansion. The power blackout of 2003 and the conclusions reached by the North American Electric Reliability Council (NERC) indicate that the nation is fast approaching a crisis stage with respect to the reliability of the transmission grids.

The transmission grid has become a "superhighway" for electric utilities to buy and sell power. Congestion from the increased flow of electricity over great distances is now a reality. The Federal Energy Regulatory Commission (FERC) has identified causes of congestion, including:

- insufficient transmission lines to match electricity generated
- inadequate transmission capacity to meet demand

According to the Edison Electric Institute:

"Between 1979 and 1989, transmission capacity grew at a slightly faster rate than the demand for electricity during peak periods. But in the subsequent years, infrastructure needs did not keep up with that demand. To handle the requirements that the transmission system expects over the next 10 years, about 27,000 GW–mi are required; however, only 6000 GW–mi are planned."

CONCLUSION

The AEE has played a vital role in developing the energy engineering profession. Through training and certification programs, professionals have gained new tools to improve the efficiency of buildings and industry.

Energy engineering is a growing profession. The AEE has played a key role in the development of the energy engineering profession and has helped professionals reach their full potential. Through the Foundation of AEE, $480,000 in scholarships have been awarded to help students in the field of energy engineering. The future of energy engineering is exceedingly bright as energy engineers seek out solutions to reduce greenhouse gases.

Climate Policy: International

Nathan E. Hultman
Science, Technology, and International Affairs, Georgetown University, Washington, D.C., U.S.A.

Abstract
Climate change is a long-term problem. Policy to address climate change faces the challenge of motivating collective action on global public goods in a world with no single international authority. International agreements can nevertheless aim to (1) reduce greenhouse gas emissions, possibly through an international emissions trading system (ETS) like that outlined in the Kyoto Protocol; (2) develop new low-emissions technology by providing incentives for cooperation on technology research and implementation; (3) provide adaptation assistance to countries and populations least able to cope with expected changes in climate.

INTRODUCTION

Anthropogenic climate change presents one of society's most vexing policy challenges. Because the problem stems largely from the fossil-fueled global economy, the costs of reducing emissions would accrue immediately and are easily quantified. On the other hand, the costs of potential damages are difficult to estimate and will likely be long-term. In addition, governing the global atmospheric commons requires a large number of actors to agree on and comply with a mechanism of self-restraint; otherwise, even the relatively virtuous would tire of the rest of the world's free-riding. Moreover, understanding the complex risks and uncertainties of climate change is a challenge even for specialists. Communicating this information accurately and effectively to a marginally interested public is harder still.

Large, long-term changes would be required to reduce the risks of climate change. While the absolute amount of greenhouse gases (GHGs) in the atmosphere affects the degree of climate change, the only variable that society can easily control is the rate of GHG emissions. To simply stabilize the absolute amount of atmospheric GHGs, the emission rate must drop to about one-half to one-third of its present levels. Deciding on what concentrations constitute moderately safe levels is challenging, and requires deriving impact estimates (such as temperature change) from possible GHG stabilization levels,[1] as well as from the projected rate of emissions reduction.

An effective international regime to govern climate change policies must balance climate protection goals with the limited enforcement ability inherent in international law.[2–4] Because the atmosphere is a common resource, protecting the climate is in most people's best interest, but no country will want to burden itself unreasonably burden itself with the excess costs of a climate friendly policy unless it believes that other countries are making equivalent sacrifices. An effective climate regime must therefore minimize free-riding. It must also ensure that the participants are complying with their obligations, which requires systems of monitoring and enforcement. Finally, because of our evolving understanding of the science of climate change and its relationship to human societies, a sound policy must remain flexible enough to incorporate new information as it becomes available, with procedures for regular scientific re-evaluation and regular review of the adequacy of the policy.

Keywords: Kyoto protocol; Climate change; Emissions trading; Clean development mechanism; Greenhouse gas; Framework convention on climate change; Global warming.

CLIMATE POLICY OPTIONS

Addressing climate change requires coordination of domestic and international systems to reduce GHG emissions and assist countries in adapting to climate change. International treaties can set guidelines for action[5,6] that are then implemented via domestic legislation.

Reducing Emissions

Several policy mechanisms can address the free-ridership and overexploitation associated with public goods like the climate. Governments can regulate the common resource directly or stipulate specific technological approaches. This command-and-control approach is relatively simple in that the rules can be set by panels of experts, and it has the appearance of fairness, because everybody must attain the same goals. However, in a diverse economy with differing costs of pollution abatement to firms, it can lead to large imbalances in the cost of compliance.

Alternatively, a governmentally set price on the externality could allow producers more flexibility. This price can be set directly as a tax (for example, $10 per ton of carbon dioxide (CO_2)), or indirectly by setting a total emissions limit and allowing entities to trade the rights to emit. These methods—taxes, emissions trading, or a hybrid of the two—can greatly reduce the total cost of compliance with the environmental target. Emissions trading systems come in two forms: in a baseline-and-credit (or permit) system, in which individual projects that result in a reduction of emissions below a pre-agreed baseline are granted credits that can be sold to firms that are not able to meet their reduction obligations. Alternatively, a cap-and-trade system sets an overall cap on emissions and then distributes, free or at auction, the entire amount of emissions allowances out to the producers.

Governments may also implement other policies to address market failures, for example by establishing minimum standards of efficiency or performance, supporting research and development of less-polluting technologies, or even guaranteeing a market for new technologies.

Equity Questions

One contentious question in developing a GHG trading system is how to allocate emissions quotas. Until now, all countries have had free access to the atmosphere; setting limits will inevitably lead to argument about who deserves a bigger slice. Three options for allocating these rights illustrate the policy challenge. The first method, usually called grandfathering, allocates permits according to what various countries or industries have emitted in the recent past. This method causes the minimum economic disruption, but it may also reward inefficient resource use and ignore the benefits that have already accrued to polluters. Alternatively, if one views the atmosphere as a universal resource or a life-support commons, then the quota may be allocated on a per-capita basis so that each person is assigned the same right to using the resource. If, on the other hand, one views the atmosphere as an economic input, the quota might be allocated to each country in proportion to its gross domestic product (GDP). The United States, for example, currently produces 24% of the world's GHG emissions,[7] has 5% of the world's population, and accounts for 21% of the world's GDP.

These simple formulae will likely not be used directly, but they do provide bases for negotiating commitments in the international community. One particularly large divide between developing and developed country positions is how to account for the cumulative GHGs emitted since the beginning of the Industrial Revolution by developed countries.[8] This atmospheric debt represents about 80% of the total anthropogenic GHG contribution, and less-developed countries' contribution will likely not equal that of developed countries until around 2100. These countries argue that richer countries should therefore move first to forestall further emissions. Developed countries often view the situation differently, pointing out that less-developed countries as a group, including China and India, will by 2010 emit GHGs at an annual rate equal to that of the developed world. Indeed China itself will soon be the world's largest GHG emitter, surpassing even the United States.

Whether generated through capture, biological sequestration, or mitigation, trading emissions requires measurement.[9] The most reliable statistics on GHG emissions relate to the burning of fossil fuels. Most countries, especially the industrialized countries that emit the most, keep detailed records of fossil fuel stocks and flows. Therefore, national emissions from fossil fuels are relatively precise and have a high amount of certainty.

Adaptation

Finally, given that some climate change is at this point inevitable, climate policy must encompass not only policies to reduce emissions of GHGs but also policies to enhance the resilience of countries to the expected changes. Often called adaptation measures, such activities include support for diversifying economies away from vulnerable crops, enhancing physical infrastructure, and bolstering institutional capabilities. Unlike policies focusing on GHGs, moreover, the benefits of adaptation accrue relatively quickly as they can immediately reduce suffering from hurricanes or floods regardless of the cause of these events.[10,11] Discussions about adaptation are often linked to questions about liability for climate damages.

EARLY INTERNATIONAL RESPONSE

Although Svante Arrhenius had postulated the existence of the greenhouse effect in 1896,[12] and significant scientific inquiry re-emerged in the 1950s, public concern about anthropogenic climate change was not significant until the late 1980s. Along with other simultaneously emerging global environmental problems like stratospheric ozone depletion and biodiversity loss, climate change moved quickly into the international arena.[13]

The international community's first concrete response was to refer the scientific questions to the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP). In 1988, the WMO and UNEP established the Intergovernmental Panel on Climate Change (IPCC), which has since become the major international expert advisory body on climate change.[14] The IPCC divides thousands of experts into three working groups on climate science, impacts of climate change, and human dimensions. It produces

Region	Greenhouse Gas Emissions billion tons CO_2e per year		
	1990	2000	KP Target
World	21.81	23.63	
Developing Countries	6.91	9.64	
Annex I	14.90	13.99	13.46
European Union	3.33	3.28	2.76
United States	4.98	5.76	4.55
Non-EU, Non-US OECD	1.84	2.20	2.05
Russia and Eastern Europe	4.75	2.74	4.19

Fig. 1 Historical emissions and Kyoto Protocol target emissions. Historical emissions from the United States Department of Energy (see Ref. 19) do not include land-use change emissions; Kyoto targets are based on net emissions reported to the UNFCCC and include land-use change emissions. Country-specific targets are available in the Kyoto Protocol text and from the UNFCCC Secretariat From UNFCC Secretariat Internet Resources (see Ref. 15) Developing Countries are defined as countries not included in Annex I.
Source: From Ref. 15.

comprehensive Assessment Reports every 5–6 years that describe the current state of expert understanding on climate, as well as smaller, targeted reports when they are requested by the international community.

THE U.N. FRAMEWORK CONVENTION ON CLIMATE CHANGE

The first international treaty to address climate change was the United Nations Framework Convention on Climate Change (UNFCCC), which entered into force in 1994 and has been ratified by 186 countries, including the United States.[15] Having emerged from the 1992 U.N. Conference on Environment and Development (the "Earth Summit"), the UNFCCC sets broad objectives and areas of cooperation for signatories. As the objective, it states that Parties to the Convention should cooperate to "prevent dangerous anthropogenic interference with the climate system." Here, dangerous is not defined explicitly but is required to include ecosystems, food supply, and sustainable economic development.

The UNFCCC identifies several important principles for guiding future treaty agreements. First, it endorses international equity. Second, it states that all signatories share a "common but differentiated responsibility" to address climate change. All countries must therefore participate, but they are allowed to do so in a way that depends on their domestic situation and historic GHG contributions. Third, the UNFCCC instructs the Parties to apply precaution in cases risking "serious or irreversible damage."

The UNFCCC also defined some emissions reduction goals for richer countries. Specifically, it grouped most developed countries into Annex I Parties (Annex I is a designation in the UNFCCC and reproduced in the Kyoto Protocol as Annex B. Annex I countries are: Australia, Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Italy, Japan, Latvia, Liechtenstein, Lithuania, Luxembourg, Monaco, Netherlands, New Zealand, Norway, Poland, Portugal, Russian Federation, Slovakia, Slovenia, Spain, Sweden, Switzerland, U.K., Ukraine, U.S.A.) and urged them to stabilize their total emissions of GHGs at 1990 levels by the year 2000. These targets were non-binding, and in retrospect, most countries did not meet these initial goals. Finally, and most importantly, the Framework Convention established a system of national emissions reporting and regular meetings of Parties with the goal of creating subsequent, more significant commitments.[16,17]

Interim Negotiations

Subsequent debates focused, therefore, on negotiating a new treaty (called a Protocol) that could enhance international action. Yet, contentious debate arose over whether developing countries would be required to agree to any reductions in return for caps on Annex I emissions.

Several negotiating blocs were solidified during this period and remain active today. The broadest split between developed and developing countries was already evident in the UNFCCC. Within developing countries, the strongest advocates for action emerged in the Alliance of Small Island States (AOSIS), an association of low-lying coastal countries around the world that are extremely vulnerable to inundation due to sea-level rise. On the other hand, the Organization of Petroleum Exporting Countries (OPEC) has been reluctant to endorse any regulation of their primary export, fossil fuel, which when burned creates the GHG CO_2.

Developed countries also have several blocs: The European Union (EU), which functions as a single legal party to the convention, has tended to favor strong action on climate change, whereas the United States, Japan, Canada, New Zealand, and Australia have been more circumspect. Russia was never an enthusiastic advocate of action on climate, but the collapse of its economy in the 1990s means that its emissions decreased considerably, allowing it some flexibility in negotiating targets. A 1995 agreement (the Berlin Mandate) adopted by the Parties to the UNFCCC stated that developing countries should be exempt from any binding commitments, including caps on their emissions, in the first commitment period. The U.S. Senate disagreed and, in 1997, declared they would not ratify any Protocol to the UNFCCC that did not call for concrete targets from developing countries.

THE KYOTO PROTOCOL

After two years of preliminary negotiations, delegates to the UNFCCC met in Kyoto, Japan in 1997 to complete a more significant treaty calling for binding targets and timetables, eventually agreeing on the Kyoto Protocol to the UNFCCC. Maintaining the principle of the Berlin Mandate, delegates rejected language that required participation by developing countries, thus damping U.S. enthusiasm. Nevertheless, the Kyoto Protocol entered into force in 2005, having been ratified by EU countries, Canada, Japan, Russia, and most developing countries. The United States and Australia are currently not Parties to the Protocol. The Kyoto Protocol builds on the UNFCCC with specific and legally binding provisions.

Targets

First, it set legally binding emissions targets for richer countries. These targets oblige Annex I Parties as a group to reduce their emissions to a level 5.2% below 1990 levels by the target period (Fig. 1). This overall average reflects reductions of 8% for the EU, 7% for the United States, and 6% for Canada; as well as increases of 8% for Australia and 10% for Iceland. Russia, whose emissions had dropped significantly between 1990 and 2000 because of economic contraction, was nevertheless awarded a 0% change, effectively providing an effort-free bonus (often called hot air). The target period is defined as 2008–2012, and countries are allowed to take an average of their emissions over this period for demonstrating compliance.

In addition, an individual country's emissions are defined as a weighted sum of emissions of seven major GHGs: CO_2, methane (CH_4), nitrous oxide (N_2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), chlorofluorocarbons (CFCs), and sulfur hexafluoride (SF_6). These gases are weighted according to a quantity called global warming potential (GWP) that accounts for different heat-trapping properties and atmospheric lifetimes of the seven gases. For a 100-year time horizon, example GWPs are: 1.00 for CO_2 (by definition), 21 for CH_4, 296 for N_2O, 100–1000 for a wide variety of halocarbons, and 22,200 for SF_6.[18] The resulting sum is reported in terms of "carbon-dioxide equivalent" or CO_2e. The Protocol allows countries to calculate a net emissions level, which means they can subtract any GHGs sequestered because of, for example, expansion of forested areas.

Implementation

The Protocol encourages countries to achieve their target primarily through domestic activities, usually called policies and measures. These include improved energy efficiency, increased use of renewable energy, and switching to lower-carbon forms of fossil fuels such as natural gas. In addition, the Protocol allows countries to offset emissions if certain domestic activities serve to absorb and sequester CO_2 from the atmosphere, thus reducing their net contribution to climate change. Allowable carbon sinks projects currently include afforestation, reforestation, forest management, cropland management, grazing land management, and revegetation. Conversely, deforestation is a process that must be counted as a cause of emissions.

The Kyoto Protocol also allows countries to obtain credits from other countries. In particular, it established three market-based mechanisms to provide states with flexibility in meeting their binding emissions reduction targets: emissions trading (ET), joint implementation (JI), and the clean development mechanism (CDM). Despite the different names, these three mechanisms are actually all forms of emissions trading—they create ways to reduce the overall cost of reaching the targets outlined above by allowing lower-cost reductions to be bought and sold on the market. All traded units are denoted in tons of CO_2e.

Emissions Trading (sometimes called Allowance Trading) is a cap-and-trade system under which the Annex I parties are assigned a maximum level of emissions (see Fig. 1), known as their assigned amount. They may trade these rights to emit through a UNFCCC registry. Only developed country Parties may participate in ET. Units of ET are termed *assigned amount units* (AAUs).

Joint Implementation is a baseline-credit system that allows trading of credits arising from projects coordinated between fully developed countries and countries in eastern Europe with economies in transition. This is a so-called project-based system, under which reductions below an independently certified baseline can be sold into the market. Joint implementation units are termed emissions reduction units (ERUs).

The CDM is another baseline-credit system that allows trading of credits arising from projects in developing countries. Another project-based system, the CDM will allow only projects that contribute both to sustainable development and to climate protection. Furthermore, they must provide benefits that would not have occurred in the absence of the CDM (so-called additionality). Post-Kyoto negotiations determined that acceptable projects include those that employ renewable energy, fossil-fuel repowering, small-scale hydroelectric power, and sinks; some projects (e.g., those under 15 MW_e) are also deemed to be "small-scale" and enjoy a streamlined approval process. Projects are subject to a process of public participation. Final acceptance of project proposals rests with the CDM Executive Board, which also approves methodologies and designates operational entities—NGOs, auditors, and other private developers—that implement and verify projects. A levy on each project will fund activities that help poor countries adapt to a changing climate. Clean development mechanism units are called certified emissions reductions (CERs).

Kyoto rules allow AAUs, ERUs, and CERs to be fungible or substitutable for each other. A final unit specific to sinks, called a removal unit (RMU), will be partially separate from this pool since it cannot be banked, or held from one commitment period to the next.

Other Provisions

The allowance assignments and flexibility mechanisms are the most significant elements of the Kyoto Protocol and its associated rules. Other noteworthy commitments include minimizing impacts on developing countries—primarily through funding and technology transfer—and establishing expert teams to develop monitoring, accounting, reporting, and review procedures.

The UNFCCC, Kyoto Protocol, and associated agreements establish three multilateral funds to assist poorer countries. The Adaptation Fund, financed through a levy on CDM transactions, is designed to help countries bolster their institutional and infrastructural capacity to manage changes in climate and damages from weather events. The Climate Change Fund focuses on technology transfer and economic diversification in the energy, transportation, industry, agricultural, forestry, and waste management sectors. Finally, the Least Developed Countries Fund exists to provide additional support in adaptation activities for the very poorest countries. The latter two funds are financed by voluntary contributions.

EUROPEAN UNION EMISSIONS TRADING SYSTEM

In 2005, the EU implemented what is to date the largest operational emissions trading system (ETS) for GHGs.[20,21] The EU–ETS is a cap-and-trade system for all 25 EU member countries, and it covers approximately 12,000 installations in six sectors (electric power; oil refining; coke ovens, metal ore, and steel; cement; glass and ceramics; paper and pulp). The plan regulates only CO_2 emissions until 2007, but thereafter other GHGs will be included. About one-half of the EU's CO_2 emissions will be regulated under the EU–ETS during this first 2005–2007 phase. The initial allocation of credits is based on individual countries' plans; countries are allowed to auction up to 5% of allowances in the first phase and 10% thereafter.

Notably, the EU–ETS contains more rigorous enforcement provisions than Kyoto. Compared to international law, the EU has far greater leverage to enforce legal provisions, and the EU–ETS imposes a steep fine (€40/$t$$CO_2$) for non-compliance. The EU–ETS replaces some national-level policies to control GHGs, notably the United Kingdom's pioneering ETS and other voluntary programs.[22] Accordingly, some facilities that had previously taken action to comply with pre-existing national laws are allowed limited exceptions to the EU–ETS.

Through a linking directive,[23] credits generated through Kyoto Protocol CDM or JI projects may be used to fulfill obligations under the EU–ETS, thereby providing an important market for these offsets. However, because of European concerns about the possible negative consequences of biological carbon sequestration (sinks) projects, CDM or JI credits generated within these categories are ineligible for the EU–ETS. The EU–ETS will thus form, by far, the largest trading program in the world and will likely set the standards for subsequent programs in other countries.

VOLUNTARY AND REGIONAL PROGRAMS

Trading programs outside national government legislation have also emerged. British Petroleum was the earliest major corporate adopter of an internal GHG trading system and, subsequently, has had a large consultative role in drafting both the Kyoto and U.K. emissions trading rules. Although the United States has declared its intention to ignore Kyoto,[24] many large American corporations have also adopted internal targets. The Chicago Climate Exchange has organized voluntary commitments from companies in the hope of establishing a position as the dominant American exchange. Yet voluntary programs, whether they derive from corporate or governmental initiatives, are ultimately constrained: often the most egregious polluters choose simply not to volunteer, and those firms that do participate may still not reduce emissions to socially desirable levels.

Many U.S. states, such as California and Oregon, have also passed or are considering legislation that would curb emissions directly or indirectly. While independent state initiatives are valuable in providing domestic innovation for the United States,[25] they are unlikely to add up to a significant reduction in global emissions. In addition, many of the companies most vulnerable to GHG regulation are asking for some guidance on what they can expect from regulators, and a state-by-state patchwork can never replace federal legislation for regulatory certainty.

CONCLUSION

International climate policy faces a period of uncertainty, innovation, and evolution over the next decade. The Kyoto Protocol remains the primary international agreement for addressing climate change globally, despite its imperfections and the continued absence of the United States and Australia. Yet, since Kyoto expires in 2012, attention has turned to negotiating a subsequent agreement that could re-engage the United States, involve China, India, and other developing countries more directly, and address concerns about compliance and enforcement.[26–28] Given the difficulties in forging an immediate, broad international consensus, the EU–ETS will likely foster the most institutional innovation and GHG market development in the near term.

The most likely interim solution, therefore, will consist of multiple, overlapping regimes that link domestic-level emissions reductions into one or more international markets.[29] In this model, the United States could, for example, institute a unilateral domestic program that addresses emissions and then allow Kyoto credits to be admissible for compliance as the EU has done. Additional agreements governing, for example, technology standards or adaptation policy may also emerge. From the perspective of energy, this evolving climate change policy will impose a carbon constraint on energy use,[30,31] most likely through a non-zero cost for GHG emissions.

ACKNOWLEDGMENTS

The author gratefully thanks the two anonymous reviewers for their thorough, detailed, and constructive comments on a previous draft.

REFERENCES

1. Mastrandrea, M.D.; Schneider, S.H. Probabilistic integrated assessment of “dangerous” climate change. Science **2004**, *304* (5670), 571–575.
2. Barrett, S. *Environment and Statecraft: The Strategy of Environmental Treaty-Making*; Oxford University Press: New York, 2003; 360.
3. Brown Weiss, E.; Jacobson, H.K. Getting countries to comply with international agreements. Environment **1999**, *41* (6), 16–20, see also 37–45.
4. Bodansky, D. International law and the design of a climate change regime. In *International Relations and Global Climate Change*; Luterbacher, U., Sprinz, D.F., Eds., MIT Press: Cambridge, Mass, 2001; 201–220.
5. Aldy, J.E.; Barrett, S.; Stavins, R.N. Thirteen plus one: a comparison of global climate policy architectures. Climate Policy **2003**, *3* (4), 373–397.
6. Sandalow, D.B.; Bowles, I.A. Fundamentals of treaty-making on climate change. Science **2001**, *292* (5523), 1839–1840.
7. United States Energy Information Administration. *Emissions of Greenhouse Gases in the United States 2003*, United States Department of Energy: Washington, DC, 2004; 108.
8. Smith, K.R. Allocating responsibility for global warming: the natural debt index. Ambio **1991**, *20* (2), 95–96.
9. Vine, E.; Kats, G.; Sathaye, J.; Joshi, H. International greenhouse gas trading programs: a discussion of measurement and accounting issues. Energy Policy **2003**, *31* (3), 211–224.
10. Pielke, R.A. Rethinking the role of adaptation in climate policy. Global Environ. Change **1998**, *8* (2), 159–170.
11. Wilbanks, T.J.; Kane, S.M.; Leiby, P.N.; Perlack, R.D. Possible responses to global climate change: integrating mitigation and adaptation. Environment **2003**, *45* (5), 28.
12. Rodhe, H.; Charlson, R.; Crawford, E. Svante Arrhenius and the greenhouse effect. Ambio **1997**, *26* (1), 2–5.
13. Bodansky, D. The history of the global climate change regime. In *International Relations and Global Climate Change*; Luterbacher, U., Sprinz, D.F., Eds., MIT Press: Cambridge, Mass, 2001; 23–40.
14. Intergovernmental Panel on Climate Change Secretariat. Sixteen Years of Scientific Assessment in Support of the Climate Convention: Geneva, 2004; 14.
15. United Nations Framework Convention on Climate Change Secretariat. UNFCCC Secretariat Internet Resources: Bonn, Germany, 2005.
16. Molitor, M.R. The United Nations climate change agreements. In *The Global Environment: Institutions, Law, and Policy*; Vig, N.J., Axelrod, R.S., Eds., CQ Press: Washington, DC, 1999; 135–210.
17. Schneider, S.H. Kyoto protocol: the unfinished agenda–An editorial essay. Climatic Change **1998**, *39* (1), 1–21.
18. Intergovernmental panel on climate change In *Climate Change 2001: The Scientific Basis*; Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A., Eds., Cambridge University Press: Cambridge, UK, 2001; 881.
19. United States Energy Information Administration *World Carbon Dioxide Emissions from the Consumption and Flaring of Fossil Fuels, 1980–2000*; United States Department of Energy: Washington, DC, 2002.
20. European Parliament. Directive 2003/87/EC Establishing a Scheme for GHG Emission Allowance Trading within the EU: Brussels, 2003.
21. Pew Center on Global Climate Change *The European Union Emissions Trading Scheme (EU–ETS): Insights and Opportunities*; The Pew Center on Global Climate Change: Washington, DC, 2005; 20.
22. Lecocq, F. *State and Trends of the Carbon Market*; Development Economics Research Group, World Bank: Washington, DC, 2004; 31.
23. European Parliament. The Kyoto Protocol’s Project Mechanisms: Amendment of Directive 2003/87/EC: Brussels, 2004.
24. Victor, D.G. *Climate Change: Debating America’s Policy Options*; Council on Foreign Relations: New York, 2004; 165.
25. Rabe, B.G. *Statehouse and Greenhouse: The Emerging Politics of American Climate Change Policy*; Brookings Institution Press: Washington, DC, 2004; 212.
26. Stavins, R.N. Forging a more effective global climate treaty. Environment **2004**, 23–30.
27. Beyond Kyoto, B.J. Foreign Affairs **2004**, *83* (4), 20–32.
28. Bodansky, D.; Chou, S.; Jorge-Tresolini, C. *International Climate Efforts Beyond 2012: A Survey of Approaches*; Pew Center on Global Climate Change: Washington, DC, 2004; 62.
29. Stewart, R.B.; Wiener, J.B. Practical climate change policy. Issues Sci. Technol. **2003**, *20* (2), 71–78.
30. Pacala, S.; Socolow, R. Stabilization wedges: solving the climate problem for the next 50 years with current technologies. Science **2004**, *305* (5686), 968.
31. Wirth, T.E.; Gray, C.B.; Podesta, J.D. The future of energy policy. Foreign Affairs **2003**, 132–155.

Coal Production in the U.S.

Richard F. Bonskowski

Fred Freme

William D. Watson

Office of Coal, Nuclear, Electric and Alternate Fuels, Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.

Abstract

U.S. coal production is historically important. Coal is now used principally as a fuel for electricity generation and other industrial processes. Coal mining practices and technologies have become more productive due to geographic changes in mining areas and increased mine sizes that produce coal at competitive prices. Coal prices respond to occasional surges in demand and to regulatory requirements but, over the long term, they have trended lower in real dollars.

INTRODUCTION

Coal has been an energy source for hundreds of years in the United States. It has helped provide many basic needs, from energy for domestic heating and cooking; to transportation for people, products, and raw materials; to energy for industrial applications and electricity generation. America's economic progress historically is linked to the use of coal from its abundant coal resources.

MILESTONES IN U.S. COAL PRODUCTION

Coal production in the United States grew steadily from the colonial period, fed the Industrial Revolution, and supplied industrial and transportation fuel during the two World Wars (Fig. 1). In 1950 the five major sectors were industrial, residential and commercial, metallurgical, electric power, and transportation with each sector accounting for 5%–25% of total consumption. From the end of World War II to 1960, coal use for rail and water transportation and for space heating declined. Coal demand grew, however, with the postwar growth in American industry and increased electricity generation starting in the early 1960s.

In 1950, U.S. coal production was 560 million short tons (mmst). In 2003, U.S. coal production was 1.07 billion short tons, an average annual increase in coal production of 1.2% per year (Table 1). Of the coal ranks in Table 1, bituminous is relatively high-Btu coal mined mostly in the East and Midwest, subbituminous is medium-Btu coal mined only in the western states, lignite is low-Btu coal principally mined in the Gulf Coast and North Dakota, and anthracite is relatively high-Btu coal mined in small quantities in Pennsylvania. With the growing importance of lower-Btu coals in the production mix over time, the energy content of coal production has not grown as rapidly as its tonnage. To depict general trends, yet allow space for selected details, coal statistics in Table 1 are shown for each year from 1993 to 2003, for every 5 years from 1953 to 1993, and for 1950. A large proportion of U.S. production is consumed domestically, so yearly coal consumption levels track coal production.

In 1973, the Arab oil embargo renewed interest in the vast U.S. coal reserves, as the nation strived to achieve energy independence. The number of coal mines and new mining capacity burgeoned. Between 1973 and 1976, coal production increased by 14.4%, or 86.3 mmst.[1] In 1978, the Power Plant and Industrial Fuel Use Act mandated conversion of most existing oil-burning power plants to coal or natural gas. New research on coal liquefaction and gasification technologies was aimed at replacing imported petroleum and supplementing domestic gas supplies. Those high-cost projects were put on hold when crude oil prices fell several years later, making synthesized coal liquids and gases uneconomic.

The shift of coal production from traditional eastern coalfields to the western United States is the most important development affecting coal markets in the last 30 years. Thick beds of low-sulfur coal with low mining cost are extensive in the Northern Great Plains states of Wyoming, Montana, and North Dakota. Starting in the 1970s, increasingly stringent restrictions on atmospheric emissions of sulfur dioxide at power plants made this coal often the most cost-effective choice for meeting sulfur dioxide limits without the installation of expensive equipment retrofits. In a matter of a few decades, a localized western resource grew to more than half of all

Keywords: Coal production; Coal preparation; Coal price; Coal mining productivity; Technology; Surface; Underground; Employment.

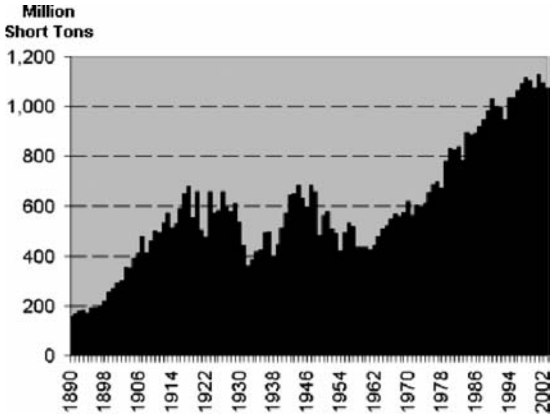


Fig. 1 United States coal production, 1890–2003.
Sources: From Energy Information Administration and Bureau of mines and U.S. geological survey (see Refs. 7 and 22).

U.S. production, from just over 60 mmst in 1973 to 549 mmst in 2003. This growth was accomplished through the deployment of long-distance coal haulage in unit trains (of more than 100 railcars moving coal only, to a single destination) and the exploitation of scale economies in the form of immense western surface coal mines. Average U.S. mine size in 2003 at 0.814 mmst per year far exceeded the average mine size in 1973 of 0.126 mmst per year. The largest U.S. mine, the North Antelope Rochelle Complex in Wyoming, alone produced over 80 mmst in 2003.

In the United States today, coal demand is driven by the electric power sector, which accounts for 90% of consumption (compared to the 19% it represented in 1950). As demand for electricity grew, demand for coal to generate it rose and resulted in increasing coal production. There were years in which coal production declined from the prior year, but, excluding years affected by a major unionized coal strike, annual increases in coal production between 1950 and 2003 outnumber decreases by almost two to one.

HOW U.S. COAL IS PRODUCED—TYPES OF MINING[2]

Growth of U.S. coal production involved expansions and adaptations in both established and evolving mining technologies. The important types of coal mining and technologies are as follows:

- Underground Mining—Extraction of coal from enclosing rock strata by tunneling below the ground surface. Also known as "deep mining," there are three types, based on mode of access. Drift mines, the easiest type to open, tunnel directly into the outcrop of horizontal or slightly sloping coal seams. Shaft mines reach the coal via a vertical shaft. In slope mines, the shaft descends on a gradient to reach the coal (suitable in hilly terrain). The following principal technologies are used in underground mining:
  - Conventional Mining—The traditional method, which employs the "room and pillar" mine layout, leaving massive pillars of undisturbed coal in place to support the overlying rock. It includes undercutting the exposed coal (the face), drilling and blasting the coal with explosive or high-pressure air, loading the broken coal into shuttle cars, and installing supplementary roof supports as needed.
  - Continuous Mining—Uses a mobile machine with forward, toothed cylinders that rotate and gouge coal from the face, where it falls onto a pan, is pulled onto loading belts and is fed to shuttle cars or movable conveyors.
  - Longwall Mining—An automated form of underground coal mining characterized by high recovery and extraction rates, feasible only in relatively flat-lying, thick, uniform coalbeds. A reserve block to be mined, the "panel," averages 1000 ft wide and 10,000 or more ft long, and is prepared by continuous mining of coal to create access tunnels on all four sides. When the longwall machinery is in place, the entire average 1000-ft width, the working face of coal is progressively sheared away, ceiling to floor, in a series of advancing passes. Dislodged coal is continuously removed via a floor-level conveyor system. Mining advances beneath automated movable roof supports within the 10,000-ft coal panel. The roof is allowed to collapse evenly in the mined out areas behind the supports.
- Surface Mining—Excavation of coal, in the most basic case, from outcroppings; or more generally, by removal of the overlying rock and soil (overburden) to expose one or more seams of coal. The following principal techniques are used in surface mining:
  - Strip Mining—An early synonym for surface mining, still widely used; to some people, it may connote irresponsible methods, without land restoration.
  - Contour Mining—A surface method used in sloping terrain, in which one or more coal beds are mined at outcrop by removing overburden to expose the coal beds.
  - Area Mining—A surface method used in flat terrain to expose coal for recovery by excavating long, successive "box cuts" or pits. Overburden excavated from the cut being mined is deposited in the previous, mined-out cut.
  - Auger and Highwall Mining—Mining usually performed within a contour or area mine, in coal in place beneath the "final highwall" (the standing

Table 1 Historical coal production by type of mining and by coal rank, selected years (production in millions of short tons)

Year	Type of mining		U.S. coal production	Bituminous coal production[b]	Subbituminous coal production	Lignite production	Anthracite production
	Underground	Surface[a]					
2003	352.8	719.0	1071.8	541.5	442.6	86.4	1.3
2002	357.4	736.9	1094.3	572.1	438.4	82.5	1.4
2001	380.6	747.1	1127.7	611.3	434.4	80.0	1.9
2000	373.7	700.0	1073.6	574.3	409.2	85.6	4.6
1999	391.8	708.6	1100.4	601.7	406.7	87.2	4.8
1998	417.7	699.8	1117.5	640.6	385.9	85.8	5.3
1997	420.7	669.3	1089.9	653.8	345.1	86.3	4.7
1996	409.8	654.0	1063.9	630.7	340.3	88.1	4.8
1995	396.2	636.7	1033.0	613.8	328.0	86.5	4.7
1994	399.1	634.4	1033.5	640.3	300.5	88.1	4.6
1993	351.1	594.4	945.4	576.7	274.9	89.5	4.3
1988	382.2	568.1	950.3	638.1	223.5	85.1	3.6
1983	300.4	481.7	782.1	568.6	151.0	58.3	4.1
1978	242.8	427.4	670.2	534.0	96.8	34.4	5.0
1973	300.1	298.5	598.6	543.5	33.9	14.3	6.8
1968	346.6	210.1	556.7	545.2	—[b]	—[b]	11.5
1963	309.0	168.2	477.2	458.9	—[b]	—[b]	18.3
1958	297.6	134.0	431.6	410.4	—[b]	—[b]	21.2
1953	367.4	120.8	488.2	457.3	—[b]	—[b]	30.9
1950	421.0	139.4	560.4	516.3	—[b]	—[b]	44.1

[a]Beginning in 2001, includes a small amount of refuse coal recovery.
[b]Subbituminous coal and lignite production were treated as bituminous coal prior to 1973 and cannot be reported separately.
Source: From Energy Information Administration (see Ref. 1).

Coal–Comm

exposed rock at the location where overburden becomes too thick for economical excavation). Auger mining uses a large-diameter drill to excavate a succession of holes within the plane of the coal bed, recovering the drilled coal. Highwall mining uses remote-controlled cutting machines, known as highwall miners—or underground mining machines, known as thin-seam miners—to mine out successive broad channels of coal from the seam left in place at the highwall.

— Mountaintop Removal (MTR) Mining—An adaptation of area mining to mountainous terrain. Often on massive scales, MTR removes all successive upper layers of rock and broad perimeters of lower rock layers. It recovers about 85%[3] of all upper coal beds contained within the rock layers and large portions of the lower beds. Mountaintop Removal operations may affect the top 250–600 ft of Appalachian peaks and ridges; they have recovered coal from as many as 18 coalbeds.[4] Mountaintop Removal mining creates huge quantities of excavated overburden that are disposed of as fill in upper portions of adjacent valleys. The fill operations are environmentally controversial, but the creation of relatively flat, developable land can be economically beneficial in steep mountainous areas.

COAL MINING TECHNOLOGY TRENDS

In the period since 1973, four distinct trends have dominated U.S. coal mining technology. The overall growth in surface coal mining at the expense of underground coal mining is the first. In 1973, underground and surface mines each accounted for 50% of total coal production. In the next 30 years, the production share from underground mines declined by a third:

Year	Underground percentage (%)	Surface percentage (%)
1973	50	50
1983	38	62
1993	37	63
2003	33	67

Growth in surface coal mining was accompanied by a second trend: the accelerated application of surface mining technology in large-scale area mines in the western region, characterized in optimal locations by box cut pits a mile or greater in length and about 200 ft wide, concentrated in the western states of Wyoming, Montana, North Dakota, Texas, Arizona, and New Mexico. In 1973, these six states accounted for 52 mmst out of a total of 599 mmst of U.S. coal mined, representing 9% of the total. By 2003, coal produced in these six western states accounted for 49% of all U.S. coal mined. No surface mines operating anywhere in the United States in 1973 had an annual output exceeding 5 mmst. By 2003, 64% of surface-mined coal was mined in the six western states in area mines exceeding 5 mmst per year of output[5]:

Surface production in western U.S. mines exceeding 5 mmst annually

Year	Percentage of U.S. total surface production (%)	Million short tons (mmst)
1973	0	0
1983	29	141
1993	50	299
2003	64	458

The third technological trend for the 1973–2003 period was the shift within underground mining from conventional room-and-pillar mining to longwall underground mining. Coal from longwall mining grew from 10 mmst in 1973 to 184 mmst in 2003, representing 52% of total U.S. underground production by 2003[6]:

Longwall production

Year	Percentage of U.S. total underground production (%)	Million short tons (mmst)
1973	3	10[a]
1983	27	80
1993	40	139
2003	52	184

[a]The 10 mmst of production in 1973 is based on 9.4 mmst of reported longwall machine coal recovery and 0.6 mmst estimated recovery by continuous mining of longwall entries.

States with substantial longwall production in 2003 included Alabama, Colorado, Pennsylvania, Utah, and West Virginia.

Due to superior productivity, large-scale surface and longwall technologies expanded faster than other mining methods. In 1983, large surface mines (greater than 5 mmst per year) had productivity higher than other surface mines, and in the next 20 years, they experienced higher rates of productivity growth. In 1983, longwall mines had about the same productivity as other underground mines; however, their productivity growth far outpaced other underground mines in the next 20 years:

Average productivity growth rates, 1983–2003

Surface Mines 5 mmst or greater	Other surface mines	Longwall mines	Other undergound mines
5.0%/year	3.1%/year	5.7%/year	2.9%/year

In the periods 1983–1993 and 1993–2003, large-scale surface technology and longwall technology saw about equal gains in productivity, decade over decade (Fig. 2). In contrast, other technologies saw decelerating gains in productivity. For a broader discussion, see the “Coal Mining Productivity” section below.

The fourth important trend was improvement in mining equipment durability and capability. Improvements to equipment, like the broad technology shifts described above, continue to raise productivity and keep coal mining costs low. For more on mining equipment, see “Mining Innovations” below.

COAL MINING PRODUCTIVITY

General production output and trends in productivity by type of coal mining by region are shown in Table 2. Effects of external and operational changes are described below. (For detailed, annual statistics, see the Energy Information Administration Web site: http://www.eia.doe.gov/emeu/aer/coal.html).

Productivity is calculated by dividing total coal production by the total direct labor hours worked by all employees engaged in production, preparation, processing, development, reclamation, repair shop, or yard work at mining operations, including office workers. In 1973, the average employee in the United States produced 2.16 short tons per hour (tph; Table 2, productivity section). In 1983, productivity had increased by 16%, to 2.50 tph; and by 1993 another 88%, to 4.70 tph. By 2003, average U.S. coal mining productivity was 6.95 short tph, an increase of 48% over 1993, and 222% over the 1973 level. Annual percentage increases in productivity during the 30-year span averaged 4.0%.

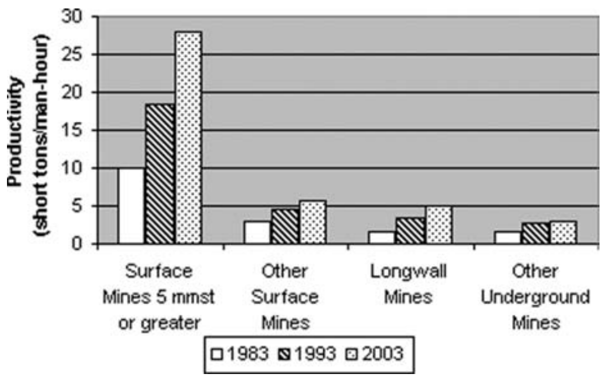


Fig. 2 U.S. Coal mining productivity.
Sources: From Energy Information Administration and Bureau of Mines and U.S. Geological Survey (see Refs. 7 and 22).

In 1973, productivity was in decline. It had fallen 10% since 1969, when the Coal Mine Safety and Health Act initiated or strengthened nationwide mine safety standards and their enforcement. This Act increased mine permitting and design requirements, added new safety and health standards in existing mines, and imposed new permitting and Black Lung fees on existing operations. The Mine Safety and Health Act of 1977 added additional safety, dust control, and mine ventilation requirements. Further, the federal government imposed strict new regulations on pollution and disruptions from mining through the Federal Water Pollution Control Act of 1972 and the Surface Mining Control and Reclamation Act of 1977. It can be argued that eventually these regulations improved productivity through safer, better-planned mines. The increasingly stringent controls of sulfur dioxide emissions under the Clean Air Act of 1970 and its amendments in 1977 and 1990 stimulated mining in low-sulfur coal regions, resulting in changes in mining techniques.

Underground coal mine productivity continued to decline through 1978, before starting a slow recovery. Underground productivity in 1973 was 1.45 tph. It fell to 1.04 tph in 1978, and then recovered to 1.61 tph by 1983. Productivity increased another 83% by 1993. By 2003, underground productivity had increased another 37%, to 4.04 tph. The annual average percentage increase in underground mining productivity for the last 30 years is nearly 4%.

Surface coal mining is less labor intensive, and its productivity is inherently higher, than underground mining. Surface productivity in 1973 was 4.56 tph. It decreased in 1983 by 16% to a level of 3.81 tph—a temporary result of the Federal Surface Mining Control and Reclamation Act of 1977, which required restoration of mined land, diverting some employees and equipment and increasing nonproduction labor hours per ton of mined coal. By 1993, surface productivity had recouped the earlier loss, increasing by 90%, to 7.23 tph. By 2003, surface productivity had increased another 49%, to 10.76 tph. The average annual percentage increase in surface mining productivity for the last 30 years is 3%.

REGIONAL PRODUCTIVITY

Regional geology, together with the type of mining, influences productivity. Appalachia (the mountainous bituminous coalfields of Pennsylvania, Maryland, Ohio, West Virginia, Virginia, eastern Kentucky, Tennessee, and Alabama, plus anthracite in Pennsylvania) has the highest number of mines, while the West has the least. As

Table 2 Production and productivity at U.S. coal mines, selected years

Item	1973	1983	1993	2003
Production(thousand short tons)				
United States	598,568	782,091[a]	945,424	1,071,753
Underground	300,080	300,379[a]	351,053	352,785
Surface	298,491	481,713[a]	594,371	718,968
Appalachian region	381,629	377,952	409,718	376,775
Underground	239,636	230,191	257,433	244,468
Surface	141,993	147,761	152,285	132,307
Interior region	156,412	173,407	167,174	146,276
Underground	56,060	49,437	56,065	52,173
Surface	100,352	123,970	111,109	94,103
Western region	60,530	225,276	368,532	548,701
Underground	10,036	18,691	37,555	56,144
Surface	50,494	206,584	330,977	492,557
Number of employees				
United States	152,204	175,642	101,322	71,023
Underground	111,799	111,888	64,604	40,123
Surface	40,405	63,754	36,718	30,900
Appalachian	124,000	126,111	71,321	46,507
Underground	96,302	90,360	50,956	30,744
Surface	27,698	35,751	20,365	15,763
Interior	22,343	34,590	18,555	11,638
Underground	12,243	16,889	10,246	6,076
Surface	10,100	17,701	8,309	5,562
Western	5,861	14,941	11,446	12,878
Underground	3,254	4,639	3,402	3,303
Surface	2,607	10,302	8,044	9,575
Number of mines				
United States	4,744	3,405	2,475	1,316
Underground	1,737	1,638	1,196	580
Surface	3,007	1,767	1,279	736
Appalachian region	4,423	2,971	2,163	1,143
Underground	1,637	1,526	1,108	521
Surface	2,786	1,445	1,055	622
Interior region	226	311	219	109
Underground	55	64	54	36
Surface	171	247	165	73
Western region	95	123	93	64
Underground	45	48	34	23
Surface	50	75	59	41
Productivity (short tons per miner-hour)				
United States	2.16	2.50	4.70	6.95
Underground	1.45	1.61	2.95	4.04
Surface	4.56	3.81	7.23	10.76
Appalachian region	1.74	1.75	3.00	3.71
Underground	1.33	1.53	2.75	3.64
Surface	3.79	2.23	3.55	3.82
Interior region	3.43	2.69	4.43	5.56
Underground	2.27	1.87	3.06	3.83
Surface	4.80	3.26	5.71	7.43

(Continued)

Table 2 Production and productivity at U.S. coal mines, selected years *(Continued)*

Item	1973	1983	1993	2003
Western region	6.64	7.60	13.53	20.82
Underground	2.59	2.28	5.23	8.42
Surface	9.64	9.63	16.49	25.01

Note: Coal-Producing Regions: Appalachian includes Alabama, eastern Kentucky, Maryland, Ohio, Pennsylvania, Tennessee, Virginia, and West Virginia; Interior includes Arkansas, Illinois, Indiana, Iowa, Kansas, western Kentucky, Louisiana, Mississippi, Missouri, Oklahoma, and Texas; Western includes Alaska, Arizona, Colorado, Montana, New Mexico, North Dakota, Utah, Washington, and Wyoming.
All statistics incorporate data for Pennsylvania anthracite coal, except for number of mines in 1973, which was not reported. Anthracite statistics were collected and published separately from other U.S. coal in 1973.
[a] Production by regions does not total production for United States, underground, and surface (first three rows). The U.S. production for 1983, by surface and underground, represent *all* mines. Details such as regional production, employment, mine count, and productivity statistics were collected that year only from mines that produced 10,000 short tons or more. The small, excluded mines represented 5.5 mmst of coal, or only 0.7 percent of total U.S. production in 1983.
Source: From Energy Information Administration (see Ref. 7).

discussed earlier, coal production and mine size grew in tandem with shifts to more surface mining and toward the West. Keen price competition motivated productivity improvements accomplished through increased mine size and production.

Appalachian productivity (Table 2) did not increase between 1973 and 1983, primarily because of productivity declines in surface mining. Between 1983 and 2003, both surface and underground mining in Appalachia improved. The annual average percentage increase in productivity in Appalachia over those 30 years was 2.6%. The backsliding in surface productivity from 1973 to 1983 corresponds with closure of more than a thousand small contour mines (many were inefficient or seasonal operations), tightening of surface mine permitting and reclamation requirements, and greater public resistance to surface mining. Some production shifted to larger surface mines and MTR operations, but their costs and workforce requirements have been relatively high in Appalachia.

Productivity in the Interior region also declined between 1973 and 1983, before picking up over the next 20 years (Table 2). Active mining in the Interior region is primarily in bituminous coals in Illinois, western Indiana, and western Kentucky and the lignite deposits of Texas, Louisiana, and Mississippi; additionally, small bituminous mines are opened from time to time in Oklahoma, Arkansas, Missouri, Kansas, and Iowa. Most surface mines in the Interior are medium or large box cut area mines. Underground mines tend to be either shaft mines or drift mines entering the coal seam beneath the final highwall. The Interior region has never supported thousands of small surface mines as had Appalachian topography. In 1983, 247 Interior region surface mines produced almost as much coal (84%) as was mined in Appalachia's 1445 mostly contour mines. From 1973 to 2003, the annual average percentage increase in surface productivity was 1.6%; in underground productivity, it was 1.8%.

The increased productivity in the West between 1973 and 1983 can all be attributed to increased surface mining. Surface productivity from 1973 to 1983 did not in itself improve, but new mines opened, and the tonnage mined from the surface quadrupled (Table 2). Those gains in surface mining production share boosted overall productivity. During that time, western underground mining, working in reserves that tended to be thick-bedded, was slightly more efficient than in thinner-bedded eastern coal. In all regions, limited longwall experience and early mine development were insufficient to significantly boost productivity or increase average mine size until after 1983. Overall, productivity of the huge surface mines, principally in Wyoming and Montana, led the productivity growth between 1983 and 2003. The average annual increase in western mine productivity from 1973 to 2003 was 3.9%. Underground productivity gains averaged 4.0% annually. Little changed in surface productivity during the 1980s and early 1990s, but by 2003, ten mines were producing more than 15 million tons per year, with great economies of scale.[7] Productivity of western surface mining improved from 9.64 short tph in 1973 to 25.01 short tph in 2003, an average annual increase of 3.2%.

CHANGES IN REGIONAL COAL PRODUCTION

The relative importance of coal production regionally has changed over the past 30 years, primarily because of the increased size and productivity of western surface mines. Production in the western region increased by more than 800% from 1973 to 2003, dominated by surface production. In 1973, western surface production was 50.5 mmst. It rose to 206.6 mmst by 1983, 331.0 mmst by 1993, and to 492.6 mmst by 2003.

Appalachian coal production varied little during the same period. Without new growth, the Appalachian region dropped from the top U.S. coal-producing region to second place, as western coal filled rising coal demand. In 1973, Appalachian production was 381.6 mmst. Production was 378.0 mmst in 1983, 409.7 mmst in 1993, and only 376.8 by 2003. The split between surface and underground mining in the Appalachian region also has been relatively stable.

Coal production in the Interior region also has changed little over the course of 30 years. Interior production was 156.4 mmst in 1973, 173.4 mmst in 1983, 167.2 mmst in 1993, and 146.3 mmst in 2003. The production split between surface and underground mines was also stable.

COAL MINING EMPLOYMENT

Employment in the coal industry from 1973 to 2003 ties in with the factors discussed above: regional coal production levels, shifts in the type of mining, and changes in productivity within regions and by mining technology. Coal mining employment includes workers at preparation plants that process the mined coal prior to sale. They are allocated to underground or surface mining proportionately based on how the coal was extracted. The average number of mine employees working daily in the United States in 1973 was 152,204 (Table 2). By 1983, daily employment had increased to a total of 175,642 (although 1983 data on employment, number of mines, and productivity covered only mines producing 10,000 short tons or more during the year). The increase resulted from surface mines hiring workers to handle reclamation and from an increase in the number of small, less efficient mines. The average number of daily employees declined to 101,322 by 1993 and to 71,023 by 2003.

Daily underground employment in 1973 averaged 111,799. That figure increased slightly by 1983, to 111,888 employees working daily. Underground employment fell to 64,604 by 1993, and to 40,123 by 2003. This trend reflects the fact that by 1983, a significant number of low productivity mines were in operation—slightly fewer underground mines produced slightly less coal using slightly more employees. Over the next 20 years, declining coal prices and increasing competition forced many of those mines out of business.

Daily surface mine employment in 1973 was 40,405. In 1983, surface employment increased to 63,754 employees, reflecting increased reclamation requirements. By 1993, surface employment was down to 36,718 employees; by 2003, it was down to 30,900.

Regional mining employment reflects the trends in production and productivity discussed above and is outlined in Table 2.

MINING INNOVATIONS

The notable improvements in mining equipment from 1983 to 2003 include the following:

- Bigger and stronger longwall face coal belt conveyors.
- Conversion to belt conveyors to move coal out of underground mines.
- Better roof-bolting equipment, (Roof bolting is a technique to secure an underground mine roof and avoid rock falls by drilling 4–12 ft up into the overlying rock layers and inserting high-tensile-strength bolts and support plates to bind together weak layers with strong layers. Bolts hold via mechanical anchors, epoxy resin, or grout.) including combination continuous-miner/bolters.
- More powerful and durable longwall cutting bits.
- Better sensors for and automation of longwall roof shields.
- More powerful and more durable electric drive motors used in many applications.
- Continuous scale-up of haul trucks, loaders, and excavators for surface mining.

A feature of the improvements listed above is that significant benefits resulted from advances in materials and technology applied to existing mining techniques, not from pioneering entirely new mining machinery. That process continues. Roof bolting was a seminal change in underground coal mining. It allowed passageways to be secured with substantially fewer timbers and “cribs” (the pillars constructed of stacked short beams used to shore up million-pound roof loads). Roof bolting—a safety standard, mandated in the 1969 Coal Mine Safety and Health Act—resulted in safer, more open mine passages and led to single-operator roof-bolting machinery far more productive than the previous labor-intensive manual timbering and cribbing.[8] For areas subject to tangential forces, steel cable roof bolts, with higher tensile strength and resistance to shear failure, give superior results. Those same qualities, along with new flexible, sprayed rock coatings, are expected to attract more proponents as mines go deeper. Though cable bolts and coatings add cost, some mines have found that fewer are needed per unit area.[9]

Examples of other recent improvements in longwall mining include variable slip clutches in the power drives for coal face belt conveyors to accommodate surges in power demand due to irregular loading and to drag from oversized coal. Ceramic facings on belt drums now give better traction and wear. Stronger materials are being marketed in roof shields to extend usable life and reduce maintenance costs. With automated operation of longwall face-shearing drums and roof-shield positioning, operators can now monitor and control mining remotely from “outby” passages at the ends of panel cuts and away from some of the noise and moving machinery. New roof bolters are highly automated and shield the operator. Advances since the 1980s in distancing the operator from the working coal face also came with the accelerated use of highwall miners, which employ video or sensor-aided monitors to give the operator effective remote control of mining for distances approaching 1000 ft. Robotic mining is expected to grow. Scaled-down longwall machines are now being tested thinner than 4 ft, and robotic cutting tools have been bench tested that can extract coal as thin as 6 in.

Computerized control systems now monitor and coordinate belt speeds in some mines from the longwall face through the face belt conveyors and all downstream belt systems, out the portal, and to storage piles. Mine operators are adopting new machinery to reduce the downtime when a longwall system is moved to a new panel. With specially designed trailers that can haul several roof shields at 13–15 MPH vs the 2–3 MPH for single shields on the common fork-lift type shield mover, one operator recently cut 2 days out of a 14-day move.[10]

Underground mines of all kinds can now take advantage of manufactured crib materials that are stronger under load than timber and more impervious to water and oil, that interlock for greater solidity, and that can be assembled quickly by machine, reducing the injury potential of personnel handling heavy materials underground. Similarly, corrugated and "pumpable" supports save time and are safer to use. Portable roof shields are now available to improve safety and coal extraction in room and pillar "retreat mining," when piers of coal that had supported the roof are removed in final mining stages.

In surface mining, new improvements include innovative use of computerized process control, which are currently being used at progressive operations along with global positioning systems to schedule and dispatch haul trucks, and to control positioning and depths of cuts by bulldozers and scrapers preparing pits and exposing the coal seam. Sensors on dozer or loader blades are guiding operators in distinguishing and recovering coal vs black shales at a growing number of mines.

The opportunities for larger surface mining equipment do have physical and practical limitations, including wheel size and tire construction, but new configurations for haul trucks are on the drawing boards that may produce a 1000-ton haul truck by 2020. At the same time, in-pit excavators will be increasing bucket size from 50 to 150 cubic yards.[11]

COAL PREPARATION

Coal preparation is processing of run-of-mine coal—the raw coal coming out of the mine—in order to enhance its characteristics for shipping and ultimate utilization. Benefits of coal preparation may include removal of noncombustible material, whose weight raises shipping costs and which can increase wear in coal grinding equipment and boilers; enhancement of deliverable heat content; removal of unwanted minerals that can foul boilers or damage the environment if entrained in boiler emissions or ash; suppression of dust; and improvement of handling and shipping qualities. These processes are carried out at preparation plants—also known as "prep" plants or wash plants—which may be located either at coal mines or at separate facilities serving numerous associated or independent mines in a mining region.

Preparation begins with crushing and screening freshly mined coal, which normally results in removal of some of the noncoal material. Some coal, especially coal from thick-bedded surface mines, is merely crushed and screened before shipping. Additional cleaning, known as mechanical cleaning, may entail separating out noncoal material in a liquid medium, which led to the widely used term "washing." The washing medium is an aqueous chemical solution prepared to enhance wettability and dissociation of the coal and noncoal materials or to produce specific gravities calibrated higher than water alone. The liquid medium may be combined with finely ground heavier minerals such as magnetite in a dense medium fluid, better to effect separation of unwanted rock and mineral matter from coal particles. Wet or "hydraulic" cleaning techniques may also include particle agitation by aeration of the coal-liquid feed, materials sorting via relative density in hydro-cyclonic chambers, and froth flotation to capture fine coal particles. To meet environmental regulations, technically advanced wash plants can remove as much as 40% of the inorganic sulfur in coal. Dry techniques, rarely used alone, include prewash segregation by vigorous shaking and pneumatic air-flow separation for crushed feed coal.

Prepared coal is commonly dewatered to some degree because excess moisture degrades deliverable heat content in the coal, and the added weight increases handling and shipping costs. Dewatering techniques range from inexpensive vibrating screens, filters, or centrifuges to the more costly use of heated rotary kilns or dryer units. Before burning, almost all coal for electric power and industrial boilers is either pulverized or crushed and sized. Precombustion coal washing is usually less costly than downstream options for removing ash and sulfur.

Two trends affect the amount of coal washed in the United States. First, production of western U.S. coal has outpaced the production of eastern U.S. coal. Most western coal is crushed and sized for market but rarely washed. Second, to meet environmental regulations, greater percentages of eastern coal are washed. In 1973, 28 and 69% of surface- and underground-mined coal, respectively, was washed. In addition, washed anthracite production, for which the type of mining was not identified, equated to 1% of U.S. production.[12] By 1983, the shares were at least 21 and 63%.[13,14] The term "at least" acknowledges that the 1973 statistics covered all mines with at least 1000 short tons of annual coal production; whereas the 1983 survey "supplement" covering prep plants was limited to larger mines, with at least 100,000 short tons of production.

Coal washing can produce large volumes of waste. In 2002, about 25% of the raw coal processed through preparation plants was discharged to waste ponds as "refuse," mixtures composed of shale, clay, coal, low-grade shaley coal, and preparation chemicals.[15] Like mining, coal-washing operations have undergone consolidation. Over the period 1983–2003, the number of U.S. wash plants fell from 362 to 132, and employment dropped from 7300 to about 2500 employees.

COAL PRICES

Except for price inflation generated following the energy crisis of 1973, U.S. coal prices were relatively stable from 1973 to 2003. When coal prices did rise, external factors like the 1973 oil embargo or burgeoning demand for coal and oil in China in 2003–2004 have been largely responsible. When real coal prices declined, however, as they did from 1975 to 2000, it was largely owing to improved labor productivity. That trend reflects the effects of "marked shifts in coal production to regions with high levels of productivity, the exit of less productive mines, and productivity improvements in each region resulting from improved technology, better planning and management, and improved labor relations."[16]

Adjusted for inflation, coal prices in year 2000 dollars decreased from $31.40 to $16.84 per short ton between 1950 and 2003 (see Table 3). The average price in nominal dollars went from $5.19 per short ton in 1950 to $17.85 in 2003. In energy terms—dollars per million Btu—coal has long been the lowest-cost fossil fuel. Petroleum products became more expensive than coal around the 1890s, when the first practical diesel and gasoline internal combustion engines were used in vehicles. Natural gas prices surpassed coal in 1979, in the first phase of natural gas price deregulation under the Natural Gas Policy Act of 1978. In 2003, one million Btu of coal sold for $0.87 on average, compared to $4.41 for natural gas and $4.75 for crude oil.[17]

The 1973 oil embargo spurred immediate and dramatic increases in coal prices, but in the long term it may have depressed prices through the long-lived excess productive capacity it generated. Between 1980 and 2000, coal prices remained under the overhang of excess capacity. Other suppliers often underbid contract coal prices considered reasonable, or even low, by mine operators.

Coal prices in Table 3 illustrate the changes that began in 1973. The average price in 1973 for U.S. coal was $8.59 per short ton, priced at the mine or original loading point. That price was unaffected by the oil embargo because in 1973 the federal government had not yet initiated policies to steer electricity producers away from petroleum as a fuel and to promote increased use of coal. Already, between 1968 and 1973, nominal coal prices had increased from $4.75 to $8.59 per short ton because coal producers passed through some of the increased costs of Black Lung taxes and the new mine safety regulations of the 1969 Coal Mine Safety and Health Act. Building on that beginning, historic real coal prices peaked in 1975, at $50.92 per short ton ($19.35 nominal).[18]

Coal prices are also commonly influenced by the end use. In the United States, the principal end uses are steam coal, metallurgical coal, and industrial coal. Steam coal, also known as "thermal" coal in international markets, is used to create steam or heat to power industrial processes. It is priced primarily on its deliverable heat content, and, because environmental regulations started in the1970s, its value may be rated down for high sulfur content. Any rank of coal may be used as a steam coal. The average nominal price at the mine or origin of all U.S. coal produced in 2003 was $17.85 per short ton. By comparison, the average price of all coal delivered to electric power plants—which generally accounts for 90% of U.S. production—was $25.91. The $8.06 difference is roughly the average cost of handling and transporting the coal to the final consumers. Examples of real delivered prices of steam coal for electricity production[19] appear below:

Steam coal at electric power plants

Year	Delivered real price per short ton (in year 2000 dollars)
1973	$28.29
1983	$53.66
1993	$32.34
2003	$24.44

Metallurgical or coking coal is used to produce metallurgical coke. The coke is produced in sealed, oxygen-free ovens and used in blast furnaces in standard iron smelting for steel production. Coke is made from bituminous coal (sometimes blended with up to 1% anthracite). It must be low in sulfur and must "agglomerate," or fuse, incorporating ash-forming minerals in the coal, to produce a strong, porous, and carbon-rich fuel that can support the load of iron ore in a blast furnace. Coal for metallurgical use requires more thorough cleaning than for steam uses and it is priced higher[20]:

Table 3 Historical U.S. coal prices at the mine or source, by coal rank, selected years (prices in dollars per short ton, expressed in nominal dollars and in inflation-adjusted year-2000 dollars)

	Average price U.S. coal sales		Average price of bituminous coal[a]		Average price of subbituminous coal		Average price of lignite		Average price of anthracite	
Year	Nominal ($)	Real ($)	Nominal ($)	Real ($)	Nominal ($)	Real ($)	Nominal ($)	Real ($)	Nominal ($)	Real ($)
2003	17.85	16.84	26.73	25.22	7.73	7.29	11.20	10.57	49.55	46.75
2002	17.98	17.27	26.57	25.53	7.34	7.05	11.07	10.63	47.78	45.90
2001	17.38	16.97	25.36	24.77	6.67	6.51	11.52	11.25	47.67	46.55
2000	16.78	16.78	24.15	24.15	7.12	7.12	11.41	11.41	40.90	40.90
1999	16.63	16.99	23.92	24.44	6.87	7.02	11.04	11.28	35.13	35.90
1998	17.67	18.32	24.87	25.78	6.96	7.21	11.08	11.49	42.91	44.48
1997	18.14	19.01	24.64	25.82	7.42	7.78	10.91	11.43	35.12	36.81
1996	18.50	19.71	25.17	26.82	7.87	8.39	10.92	11.64	36.78	39.19
1995	18.83	20.44	25.56	27.75	8.10	8.79	10.83	11.76	39.78	43.19
1994	19.41	21.50	25.68	28.45	8.37	9.27	10.77	11.93	36.07	39.96
1993	19.85	22.46	26.15	29.59	9.33	10.56	11.11	12.57	32.94	37.27
1988	22.07	29.16	27.66	36.54	10.45	13.81	10.06	13.29	44.16	58.34
1983	25.98	39.84	31.11	47.71	13.03	19.98	9.91	15.20	52.29	80.19
1978	21.86	47.77	22.64	49.48	—[a]	—[a]	5.68	12.41	35.25	77.04
1973	8.59	26.97	8.71	27.35	—[a]	—[a]	2.09	6.56	13.65	42.86
1968	4.75	19.07	4.70	18.87	—[a]	—[a]	1.79	7.19	8.78	35.24
1963	4.55	20.87	4.40	20.19	—[a]	—[a]	2.17	9.96	8.64	39.64
1958	5.07	24.73	4.87	23.76	—[a]	—[a]	2.35	11.46	9.14	44.59
1953	5.23	28.67	4.94	27.08	—[a]	—[a]	2.38	13.05	9.87	54.10
1950	5.19	31.40	4.86	29.40	—[a]	—[a]	2.41	14.58	9.34	56.50

[a]Through 1978, subbituminous coal is included in "Bituminous Coal".
Source: From Energy Information Administration (see Ref. 7).

Coal–Comm

Coal at metallurgical coke plants

Year	Delivered real price per short ton (in year 2000 dollars)
1973	$62.07
1983	$90.94
1993	$53.68
2003	$47.77

Industrial coal can be of any rank. It is coal used to produce heat for steam or industrial processes. Typical industrial coal consumers include manufacturing plants, paper mills, food processors, and cement and limestone products. Prices[21] tend to be higher than for coal received at electricity producers, primarily because average tonnages purchased by industrial consumers are smaller and, in some cases, because the plant processes require specific or less-common coal characteristics:

Coal at other industrial facilities

Year	Delivered real price per short ton (in year 2000 dollars)
1973	NA
1983	$60.30
1993	$36.47
2003	$32.74

CONCLUSION

In recent years, about 90% of coal production in the United States has been consumed at domestic electric power plants. Coal use grew because of secure, abundant domestic reserves and relatively low prices. Demand has been met through increasing mine productivity, which in turn has been supported by operation of larger and larger mines, the use of larger, more efficient mining machinery, advances in technology and control systems, and the employment of fewer mine personnel.

REFERENCES

1. *Annual Energy Review 2004*, DOE/EIA-0384(2004); Energy Information Administration: Washington, DC, 2005; 207.
2. Definitions in the section are adapted from: *Coal Glossary*, Internet Website, Energy Information Administration: Washington, DC, 2005, http://www.eia.doe.gov/cneaf/coal/page/gloss.html (accessed August 2005); *A Dictionary of Mining, Mineral, and Related Terms*: U.S. Department of the Interior, Bureau of Mines: Washington, DC, 1968, various.
3. McDaniel, J.; Kitts, E. *A West Virginia Case Study*, Electronic slide presentation. Mining and Reclamation Technology Symposium: Morgantown, WV, 1999, http://www.epa.gov/region3/mtntop/appendix.htm "Case Study," (accessed August 2005).
4. *Draft Programmatic Environmental Impact Statement, Mountaintop Mining/Valley Fills in Appalachia*, http://www.epa.gov/region3/mtntop/eis.htm, Chapter 3, U.S. Environmental Protection Agency, Mid-Atlantic Region: Philadelphia, PA, 2003; III.J-1.
5. Mine Safety and Health Administration, Form 7000-2, *Quarterly Mine Employment and Coal Production Report*, file data, 1983–2003; Coal—bituminous and lignite chapter. *Minerals Yearbook 1973*; U.S. Department of the Interior, Bureau of Mines: Washington, DC; 326.
6. DOE/EIA-TR-0588. *Longwall Mining*; Energy Information Administration: Washington, DC, 1995; 34–35; Coal—bituminous and lignite chapter. *Minerals Yearbook 1973*; U.S. Department of the Interior, Bureau of Mines: Washington, DC; 353.
7. DOE/EIA-0584(2003). *Annual Coal Report*; Energy Information Administration: Washington, DC; 22; http://www.eia.doe.gov/cneaf/coal/page/acr/acr_sum.html (accessed September 2005).
8. Barczak, T. Q&A—research. Am. Longwall Mag. **2005**, *August*, 45.
9. Peterson, D.; LaTourrette, T.; Bartis, J. *New Forces at Work in Mining: Industry Views of Critical Technologies*; RAND Corporation: Arlington, VA, 2001; 26, [http://www.rand.org/publications/MR/MR1324/MR1324.ch3.pdf (accessed August 2005)].
10. Bahr, A. Saving time and money at elk creek. Am. Longwall Mag. **2005**, *August*, 43.
11. Peterson, D.; LaTourrette, T.; Bartis, J. *New Forces at Work in Mining: Industry Views of Critical Technologies*; RAND Corporation: Arlington, VA, 2001; 28, [http://www.rand.org/publications/MR/MR1324/MR1324.ch3.pdf (accessed August 2005)].
12. Coal—bituminous and lignite chapter. *Minerals Yearbook 1973*; U.S. Department of the Interior, Bureau of Mines: Washington, DC; 358–359, 391.
13. *Coal Production 1984*, Appendix C, DOE-EIA-0118(84), Energy Information Administration: Washington, DC; 118 (for quantities cleaned, by type of mining).
14. *Coal Production 1983*, DOE-EIA-0118(83), Energy Information Administration: Washington, DC; 4 (for quantities mined, by type of mining).
15. *Bituminous Coal Underground Mining: 2002*, U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2004, Tables 6a and 7.
16. Flynn, E. *Impact of Technological Change and Productivity on the Coal Market, Issues in Midterm Analysis and Forecasting*, DOE-EIA-0607(2000); Energy Information Administration: Washington, DC; 1; http://www.eia.doe.gov/oiaf/analysispaper/pdf/coal.pdf (accessed October 2005).
17. *Annual Energy Review 2004*, DOE/EIA-0384(2004), Energy Information Administration: Washington, DC, 2005; 67.
18. *Annual Energy Review 2004*, DOE/EIA-0384(2004), Energy Information Administration: Washington, DC, 2005; 219.

19. *Annual Summary of Cost and Quality of Electric Utility Plant Fuel 1976*; Bureau of Power, Federal Power Commission: Washington, DC, 1977; 12; *Cost and Quality of Fuels for Electric Utility Plants1983*, DOE/EIA-0191(83); Energy Information Administration: Washington, DC, 1984; 27; *Electric Power Monthly*, Energy Information Administration: Washington, DC, 68; http://www.eia.doe.gov/cneaf/electricity/epm/table4_1.html (accessed as Tables 4.1, October 2005).
20. *Quarterly Coal Report*, October–December 1984, 1994, and, 2004, DOE/EIA-0121(84) and (94) and (2004), Energy Information Administration: Washington, DC, 17, 53, 26; Coal—bituminous and lignite chapter. *Minerals Yearbook 1973*, U.S. Department of the Interior, Bureau of Mines: Washington, DC; 320.
21. *Quarterly Coal Report*, October-December 1984, 1994, and, 2004, DOE/EIA-0121(84) and (94) and (2004), Energy Information Administration: Washington, DC, 17, 53, 28.
22. U.S. Bureau of Mines, Annual Coal Production Surveys; Bureau of Mines and U.S. Geological Survey: Washington, DC, 1975.

BIBLIOGRAPHY

1. *Annual Coal Report*, DOE/EIA-0118 (years 1976–1992), Energy Information Administration: Washington, DC.

Coal Supply in the U.S.

Jill S. Tietjen
Technically Speaking, Inc., Greenwood Village, Colorado, U.S.A.

Abstract
The United States has approximately 275 billion tons of coal resources in the ground, enough to last more than 200 years at current rates of consumption. Coal ranks in the United States are (from lowest to highest) lignite, subbituminous, bituminous, and anthracite. Estimated recoverable reserves of coal are located in 32 states, and mining currently takes place in 26 of them. Coal is mined in underground and surface mines. Coal is transported from the mines to the ultimate users by conveyer belt, truck, barge, ship, train, or coal slurry pipeline and sometimes by multiple methods. More than 50% of the electricity generated in the United States comes from coal-fired power plants. Electric utilities and industrial users are the largest consumers of coal.

INTRODUCTION

The United States has approximately 275 billion tons of coal resources in the ground, enough to last more than 200 years at current rates of consumption. Coal constitutes 95% of the fossil energy reserves in the United States. Half of the recoverable coal reserves in the world are in China, the United States, and the former Soviet Union.[1]

These estimated coal reserves are generally characterized by coal rank. The coal ranks in the United States are (from lowest to highest) lignite, subbituminous, bituminous, and anthracite. Estimated recoverable reserves of coal are located in 32 states, and mining currently takes place in 26 of them. Coal deposits occur in additional states, but either reserve tonnages have not been estimated or physical conditions are not conducive to mining. Coal is mined in underground mines and in surface mines. Coal is transported from the mines to the ultimate users by conveyer belt, truck, barge, ship, train or coal slurry pipeline, and sometimes by multiple methods. More than 50% of the electricity generated in the United States comes from coal-fired power plants. Electric utilities and industrial users are the largest consumers of coal.[2]

RANKS OF COAL

Coal results from geologic forces having altered plant materials in different ways. The four coal types are derived from peat, which is the first stage in the formation of coal. Peat is partially decomposed plant material. The four types or ranks of coal are lignite, subbituminous, bituminous, and anthracite.[3] The rank of a coal refers to the degree of metamorphosis it has undergone. The longer the organic materials comprising the coal have been buried, along with the amounts of pressure and heat imposed, the greater is its conversion to coal.

Lignite is the lowest rank of coal and is often referred to as brown coal. It has the lowest heating value of any of the four categories of coal [4,000–8,300 Btu/lb (A British thermal unit (Btu) is the amount of heat needed to raise the temperature of 1 lb of water 1°F)]. It is brownish-black and has a very high moisture (sometimes as high as 55%) and ash content, and cannot be transported very long distances economically. About 8% of the coal produced in the United States is lignite, and most of it is in Texas and North Dakota. Lignite is primarily used as a fuel for electricity generation.

Subbituminous coal ranges from dull dark brown, soft and crumbly, to bright jet black, hard and relatively strong. Although it has a heating value higher only than lignite, averaging 8,400–8,800 Btu/lb in the Powder River Basin (PRB) of Wyoming, there are plentiful reserves of subbituminous coal in the West and in Alaska. Although this coal has a moderately high moisture content (20%–30%), the PRB coal resources in Wyoming and Montana have a lower sulfur content than many bituminous coal reserves and, thus, burn more cleanly. More than 90% of subbituminous coal production comes from the PRB. More than 40% of the coal produced in the United States is subbituminous.

Bituminous coal, often called soft coal in Europe, is the most common type of coal used for the generation of electricity in the United States. It is a dense coal, usually black and sometimes dark brown. This coal has a heating value from 10,500 to 15,000 Btu/lb. The primary use of bituminous coal—and all U.S. coal—is as "thermal" or steam coal, consumed mostly for electricity generation. Bituminous coal also has properties that allow it to be used as metallurgical coal for the steel and iron industries. Bituminous coal accounts for about one-half of U.S. coal production.

Keywords: Anthracite; Bituminous; Subbituminous; Lignite; Coal reserves; Underground mining; Surface mining; Barge; Railroad; Truck; Coal basin.

Table 1 U.S. coal regions and coal fields

Coal region	Coal field	States
Appalachia	Northern Appalachia	MD, OH, PA. Northern WV
	Central Appalachia	Eastern KY, VA, Southern WV, Northern TN
	Southern Appalachia	AL, Southern TN
Interior	Illinois Basin	Western KY, IL, IN
	Gulf Coast lignite	TX, LA, MS
	Other Western Interior	AR, IA, KS, MO, OK
West	Powder River Basin	WY, MT
	North Dakota lignite	ND
	Southwest	AZ, NM
	Rockies	CO, UT
	Northwest	AK, WA

Source: From Energy Information Administration (see Ref. 6).

Anthracite is the hardest coal (often referred to as hard coal), is brittle, has a high luster, and gives off the second greatest amount of heat when it burns (averaging 12,500 Btu/lb). It is low in volatile matter and has a high percentage of fixed carbon. Anthracite accounts for a small amount of the total coal resources in the United States. It is found mainly in Pennsylvania and is generally used for space heating. Since the 1980s, anthracite refuse or mine waste has been used to generate electricity.[4]

COAL BASINS IN THE UNITED STATES

The large majority of coal in the contiguous United States is found in the Appalachian, Interior, and Western coal regions. These regions are identified in Table 1 and Fig. 1. The characteristics of some typical kind of coals are shown in Table 2. Coal varies significantly from mine to mine and from region to region, even within the same classification (e.g., bituminous).[5]

The Appalachian coal region contains the largest deposit of high-grade bituminous coals in the United States. It is generally divided into three parts. Anthracite and bituminous coals are found in northern Appalachia, which includes the bituminous coal deposits found in the states of Pennsylvania, Maryland, Ohio, and northern West Virginia, and the anthracite fields of eastern Pennsylvania. The bituminous coals found in the central Appalachian area of southern West Virginia, Virginia, northern Tennessee, and eastern Kentucky include deposits that are low in sulfur and highly desirable as a fuel for electricity generation. The Alabama and southern Tennessee coal deposits that are characterized as southern Appalachia have been used primarily in the steel industry through history.

The Interior Basin coals, in general, contain a lower heating value than Appalachian coals, with higher sulfur

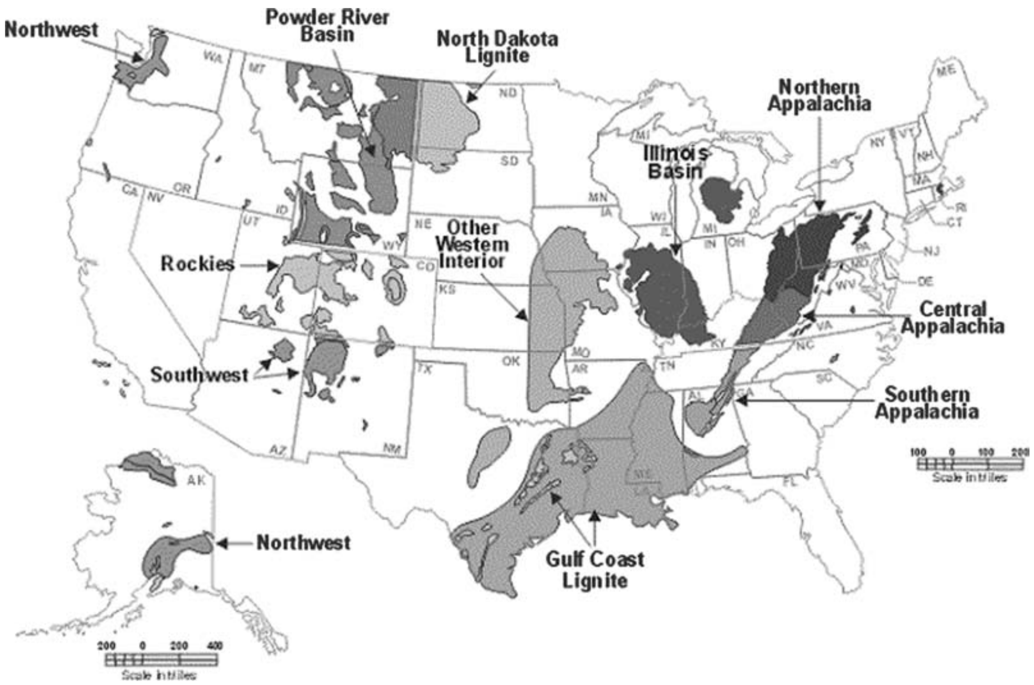


Fig. 1 Coal regions and coal fields.

Coal–Comm

Table 2 Characteristics of some typical U.S. coals

Coal region	Type of coal	Heating value (Btu/lb)	% Moisture	% Sulfur
Appalachia	Anthracite	12,440	2.4	0.5
Appalachia	Bituminous	12,790	3.3	1.0
Illinois Basin	Bituminous	11,440	8.5	4.3
Western	Bituminous	12,320	2.2	0.6
PRB	Subbituminous	8,400–8,800	28.0	0.2
Gulf Coast	Lignite	6,490	34.8	0.5

Source: From Combustion Engineering, Inc. (see Ref. 7).

content. Mining primarily is found today in the Illinois Basin states of Illinois, Indiana, and western Kentucky. In the Illinois Basin, most coal with lower chlorine content has already been mined, and most of the desirable remaining coal will need to be mined in deep underground mines in the future. Bituminous coal deposits are found in other states in the region, including Michigan, Iowa, Missouri, Nebraska, Kansas, Arkansas, Oklahoma, and part of Texas.

Lignites in the Gulf Coast region are found in Alabama, Mississippi, Louisiana, Texas, and Arkansas. Some of these lignites have a moisture content as high as 55% and heating values sometimes lower than 4,000 Btu/lb, which limits their marketability. Some of the Louisiana lignites are similar to those in North Dakota, with a lower moisture content of around 36% and sodium levels of 5%–8%.

Very large deposits of various kinds of coal are found in the Great Plains. Immense deposits of lignite are found in North and South Dakota. The PRB of Wyoming and Montana contains a very large deposit of subbituminous coal.

The Rocky Mountain states of Colorado and Utah contain deposits of bituminous coal, although other coals can be found throughout the region. Coal from the Southwest, primarily from New Mexico, is both bituminous and subbituminous, and is used for electricity generation. All four ranks of coal exist in Washington state, but the majority is subbituminous. Subbituminous coal is produced at the state's only active mine and is used for electricity generation.

Coal is distributed widely throughout Alaska, differing greatly in rank and geologic environment. Reserves are estimated to be 15% bituminous and 85% subbituminous and lignite. Developed reserves in Alaska are primarily located near the main lines of transportation.[8]

METHODS OF MINING

Coal is removed from the earth either through underground mining or surface mining. About two-thirds of current U.S. coal production comes from surface mines.

Underground mining, also referred to as deep mining, is used when the coal is more than several hundred feet underneath the surface of the earth. There are three types of underground mines: drift mines, slope mines, and shaft mines. The type of mine that will be constructed is dependent on the depth of the coal seam and the terrain. Drift mines have horizontal entries into the coal seam from a hillside. Slope mines, which are usually not very deep, are inclined from the surface to the coal seam. Shaft mines are generally the deepest type of mine and have vertical access to the coal seam via elevators that carry workers and equipment into the mine. Some underground mines are more than 2,000 ft underground. These types of mines are shown in Fig. 2.[9]

Surface mining can usually be used when the coal is buried less than 200 ft under the surface of the earth. Large machines remove the overburden—the layers of soil, rock, and other materials that are between the surface of the earth and the highest minable coal seam (see Fig. 3). Then the exposed coal seam is blasted, using explosives, and the coal is removed. Except in the PRB, where a single minable seam may be as thick as 100 ft, most surface mines produce from multiple coal seams, layered within the rock strata. When that is done, the process outlined above is repeated, removing successive layers of rock (called interburden) between mined coal and coal. For a surface mine, the ratio of overburden to the amount of coal removed is called the overburden ratio. Lower ratios mean that the mine is more productive, and the ratios may be lowered by recovering additional coal seams.

In most cases, after the coal has been mined, the materials that had been on the top of the coal are restored,

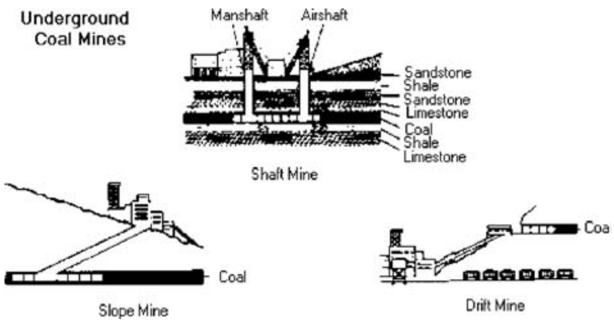


Fig. 2 Underground mining.

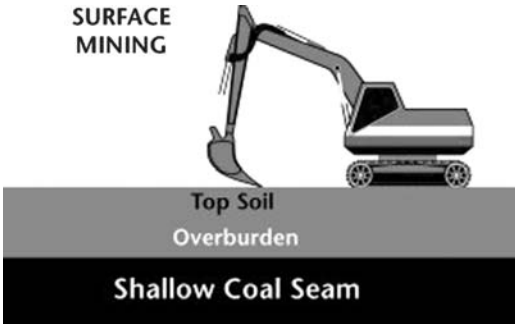


Fig. 3 Surface mining.

so that the surface on top of the pits can be reclaimed. When the area has been replanted, it can often be used for any purpose that could have occurred before the land was mined: cropland; wildlife habitat; recreation; or offices, homes, and stores. In the case of mountaintop removal mining, which is used primarily in Appalachia, the surface on top of the coal is not restored to its original contour.

There are several types of surface mines: area, contour, mountaintop removal, auger, and open pit. Area surface mines are usually found in flat terrain and consist of a series of cuts 100–200 ft wide. The overburden from one cut is used to fill in the mined-out area of the previous cut. Contour mining, which occurs in sloping and mountainous terrain, follows a coal seam along the side of a hill. Open pit mining is generally found where the coal seams are thick and the mines can reach depths of several hundred feet.[10]

COAL PRODUCTION

Coal is mined from surface and underground mines in 26 states. More coal is mined in Wyoming than in any other state. Other states with high production include West Virginia, Kentucky, Pennsylvania, and Texas. Table 3 presents coal production by state for 2001 through 2005. The total 2005 production—just over 1 billion short tons of coal—represents more than one-fifth of the world's coal production.

The Appalachian coal region annually produces about 35% of total U.S. coal production from large underground and small surface mines. West Virginia is the largest coal-producing state in the Appalachian coal region and the second largest in the United States. Coal from this region is used as a power plant fuel to produce electricity, for production of metals, and to export to other countries.

Texas is the largest coal-producing state in the Interior region, almost all lignite for power plants. The Illinois Basin states are the next more significant producers: Indiana, Illinois, and Western Kentucky, respectively. Almost 13% of U.S. coal production is from the Interior region, most produced in midsize surface mines by midsize to large companies.

Wyoming is the largest producing state in the nation, a position it has held for 19 consecutive years as of 2005. Its 2005 level of production was just 11 million tons short of the production of the next five largest coal-producing states combined (West Virginia, Kentucky, Pennsylvania, Texas, and Montana). About 52% of the coal mined in the United States comes from the Western region, with more than 30% from Wyoming alone. The large surface mines in the West are the largest coal mines in the world.[12]

COAL TRANSPORTATION

Coal in the United States moves from the coal basin in which it was mined to its final use by rail, water, truck, tramway, conveyor, or slurry pipeline, or more than one of those modes (so-called multimodal transportation). Power plants that are located near or at a mine and burn coal from that mine are called minemouth power plants. The coal for these plants usually moves from the mine to the power plant by truck, tramway, or conveyor. Coal that has been mixed with a liquid, usually water, is called a slurry and is moved through a pipeline that is called a slurry pipeline. The longest coal slurry pipeline in operation in the United States moved coal 273 mi from Arizona to Nevada until January 1, 2006. (It was closed at that time pending the completion of environmental upgrades at the power plant where it is consumed). More than 65% of coal in the United States is transported for at least part of its trip to market by train.[13]

The primary manner in which coal was transported in 2002 is shown in Table 4. Fig. 4 shows the methods of transportation graphically, demonstrating the strong predominance of rail transportation for moving coal in the United States.

USES OF COAL

The overwhelming majority of coal in the United States is used by the electric utility sector, as shown in Table 5—almost 92% of all of the coal consumed in the United

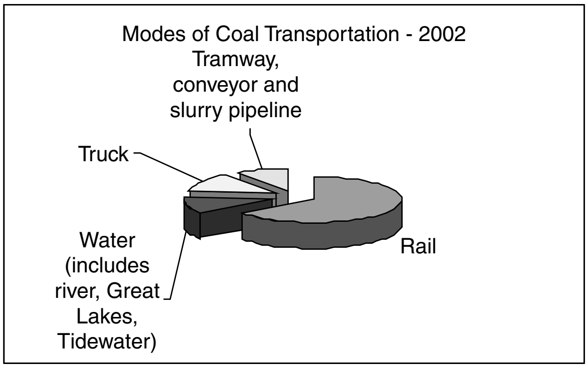


Fig. 4 Modes of coal transportation 2002.

Coal–Comm

Table 3 U.S. coal production by coal-producing region and state 2001–2005 (million short tons)

Coal producing region and state	2001	2002	2003	2004	2005
Appalachian total	**431.2**	**396.2**	**376.1**	**389.9**	**396.4**
Alabama	19.4	18.9	20.1	22.3	21.3
Kentucky, Eastern	109.1	99.4	91.3	90.9	93.4
Maryland	4.6	5.1	5.1	5.2	5.2
Ohio	25.4	21.2	22.0	23.2	24.7
Pennsylvania total	74.1	68.4	63.7	66.0	67.3
Anthracite	1.5	1.3	1.2	1.7	1.6
Bituminous	72.7	67.1	62.5	64.3	65.6
Tennessee	3.3	3.2	2.6	2.9	3.2
Virginia	32.8	30.0	31.6	31.4	27.7
West Virginia	162.4	150.1	139.7	148.0	153.6
Northern	38.2	34.0	34.9	40.6	42.6
Southern	124.5	116.0	104.8	107.3	110.9
Interior total	**146.9**	**146.6**	**146.0**	**146.0**	**149.2**
Arkansas	—	—	—	—	—
Illinois	33.8	33.3	31.6	31.9	32.1
Indiana	36.7	35.3	35.4	35.1	34.4
Kansas	0.2	0.2	0.2	0.1	0.2
Kentucky, Western	24.7	24.7	21.5	23.4	26.4
Louisiana	3.7	3.8	4.0	3.8	4.2
Mississippi	0.6	2.3	3.7	3.6	3.6
Missouri	0.4	0.2	0.5	0.6	0.6
Oklahoma	1.7	1.4	1.6	1.8	1.8
Texas	45.0	45.2	47.5	45.9	45.9
Western total	**547.9**	**550.4**	**548.7**	**575.2**	**587.0**
Alaska	1.5	1.1	1.1	1.5	1.5
Arizona	13.4	12.8	12.1	12.7	12.1
Colorado	33.4	35.1	35.8	39.9	38.5
Montana	39.1	37.4	37.0	40.0	40.4
New Mexico	29.6	28.9	26.4	27.2	28.5
North Dakota	30.5	30.8	30.8	29.9	30.0
Utah	27.0	25.3	23.1	21.7	24.5
Washington	4.6	5.8	6.2	5.7	5.3
Wyoming	368.7	373.2	376.3	396.5	406.4
Refuse recovery	**1.8**	**1.0**	**1.0**	**1.0**	**0.7**
U.S. total	**1127.7**	**1094.3**	**1071.8**	**1112.1**	**1133.3**

Source: From Energy Information Administration (see Ref. 11).

States. Industries and businesses also burn coal in their own power plants to produce electricity. The coal is burned to heat water to produce steam that turns turbines and generators to produce electricity.

Industries across the United States use coal for heat and as a chemical feedstock. Derivatives of coal, including methanol and ethylene, are used to make plastics, tar, synthetic fibers, fertilizers, and medicine. The concrete and paper industries also burn coal. Altogether, industrial customers consume more than 7% of the coal mined in the United States.

In the steel industry, coal is baked in hot furnaces to make coke, which is used to smelt iron ore into the pig iron or hot iron needed for making steel. The very high

Table 4 Coal transportation in the United States (millions of short tons, 2002)

Mode of transportation	Tonnage moved by this means
Rail	685,086
Water (includes river, Great Lakes, tidewater)	126,870
Truck	138,222
Tramway, conveyor and slurry pipeline	99,986
Total 2002	1,051,406

Source: From Energy Information Administration (see Ref. 14).

Table 5 Coal consumption by sector

Sector	2001	2002	2003	2004	2005
Electric power	964.4	977.5	1005.1	1016.3	1039.0
Coke plants	26.1	23.7	24.2	23.7	23.4
Other industrial plants	65.3	60.7	61.3	62.2	60.8
Combined heat and power (CHP)	25.8	26.2	24.8	26.6	20.6
Non-CHP	39.5	34.5	36.4	35.6	40.2
Residential and commercial users	4.4	4.4	4.2	5.1	5.1
Residential	0.5	0.5	0.5	0.6	0.6
Commercial	3.9	4.0	3.8	4.6	4.6
Total	1060.1	1066.4	1094.9	1107.3	1128.3

Source: From Energy Information Administration (see Ref. 15).

temperatures and the fluxing properties to isolate impurities made possible by using coke give steel the strength and flexibility required for bridges, buildings, and automobiles.

Coal provides more than half of the electricity in this country, as shown in Fig. 5, and in certain areas of the country accounts for about two-thirds of the fuel mix for electric power generation. Natural gas use has increased significantly in recent years, as newer generating facilities over the past decade were almost exclusively natural gas-fired because of formerly lower gas prices, lower emissions, and lower investment costs and lead times.[16] With the passage of recent legislation, including the Clean Air Interstate Rules and the Energy Policy Act of 2005, the federal government has put in place a number of incentives to encourage more use of clean, coal-fired power plants and the use of coal for synthetic natural gas and liquid transportation fuels.

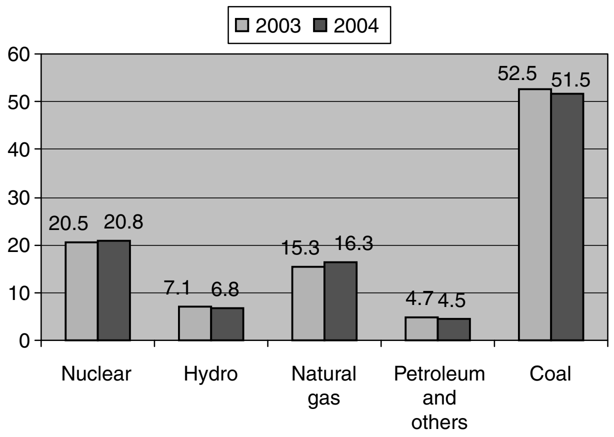


Fig. 5 Share of electric power sector net generation by energy sources 2003 and 2004.

CONCLUSION

Coal is a fossil fuel resource in the United States that serves an important role as a feedstock for providing electricity and fuel and/or feedstock for a variety of industries, including steel and plastics. Railroads move more than half of the coal from the mine to its final destination. Although natural gas usage has increased significantly for providing electricity over the past decade, coal remains the primary fuel for electricity production.

Coal–Comm

REFERENCES

1. Energy Information Administration. U.S. Coal Reserves: 1997 Update. DOE/EIA-0529(97). Available at: http://tonto.eia.doe.gov/FTPROOT/coal/052997.pdf (accessed August 2006); Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006); The Environmental Literacy Council. Coal. Available at: www.enviroliteracy.org/article.php?id=18&print=1 (accessed April 2006).
2. Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006); The Environmental Literacy Council. Coal. Available at: www.enviroliteracy.org/article.php?id=18&print=1 (accessed April 2006).
3. Utah Mining Association. Types of Coal. Available at: www.utahmining.org/coaltype.htm (accessed April 2006).
4. Energy Information Administration. Coal: Glossary. Available at: www.eia.doe.gov/cnear/coal/page/gloss.html (accessed August 2006); The Environmental Literacy Council. Coal. Available at: www.enviroliteracy.org/article.php?id=18&print=1 (accessed April 2006); Texas Railroad Commission's Kids World School House. Types of Coal. Available at: http://kids.rrc.state.tx.us/school/classroom/coal/types.html (accessed April 2006); Utah Mining Association. Types of Coal. Available at: www.utahmining.org/coaltype.htm (accessed April 2006); Powder River Coal Company. Types of Coal. Available at: www.ccsd.k12.wy.us/mines/PR/CoalTypes.html (accessed April 2006); Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006); Lignite. Available at: http://enl.wikipedia.org/wiki/Lignite (accessed April 2006); Sub-bituminous Coal. Available at: http://en.wikipedia.org/wiki/Sub-bituminous_coal (accessed April 2006); Bituminous Coal. Available at: http://en.wikipedia.org/wiki/Bituminous_coal (accessed April 2006); Anthracite Coal. Available at: http://en.wikipedia.org/wiki/Anthracite (accessed April 2006).
5. Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006).
6. Energy Information Administration. Coal Transportation: Rates and Trends—1979–2001 (partial 2002). Available at: www.eia.doe.gov/cneaf/coal/page/trans/ratesntrends.html (accessed March 2006).
7. Singer, J.G. Ed., *Combustion: Fossil Power: A Reference Book on Fuel Burning and Steam Generation*; Combustion Engineering, Inc.: Windsor, CT, 1991; Accounting for Changes in Coal Properties When Optimizing Combustion. *Lehigh Energy Update*. 2002; 20(1). Available at: www.lehigh.edu/energy/leu/march_leu_2002.pdf (accessed August 2006); Mercury in U.S. Coal. U.S. Geological Survey Open-File Report 98-0772. Available at: http://pubs.usgs.gov/of/1998/of98-772 (accessed August 2006).
8. Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006); Freme F. U.S. Coal Supply and Demand: 2004 Review. Energy Information Administration. Available at: www.eia.doe.gov/cneaf/coal/page/special/feature.html (accessed March 2006).
9. United Mine Workers of America. Underground Mining: Types of Underground Coal Mines. Available at: www.umwa.org/mining/ugtype.shtml (accessed April 2006); Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006).
10. Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006); United Mine Workers of America. Surface Coal Mining. Available at: www.umwa.org/mining/surmine.shtml (accessed April 2006); University of Kentucky and Kentucky Geological Survey. Methods of Mining. Available at: www.uky.edu/KGS/coal/coal_mining.htm (accessed August 2006).
11. Freme, F. *U.S. Coal Supply and Demand: 2005 Review*. Energy Information Administration. Available at: www.eia.doe.gov/cneaf/coal/page/special/feature.html (accessed August 2006). (The data for 2005 are preliminary).
12. Freme, F. *U.S. Coal Supply and Demand: 2004 Review*, Energy Information Administration. Available at: www.eia.doe.gov/cneaf/coal/page/special/feature.html (accessed March 2006).
13. Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006); The Center for Land Use Interpretation. Black Mesa Coal Mine and Pipeline. Available at: http://ludb.clui.org/ex/i/AZ3134 (accessed April 2006).
14. Energy Information Administration. Coal Transportation: Rates and Trends in the United States, 1979–2001 (with supplementary data to 2002). Available at: www.eia.doe.gov/cneaf/page/trans/ratesntrends.html (accessed March 2006).
15. Freme, F. *U.S. Coal Supply and Demand: 2004 Review*. Energy Information Administration. Available at: www.eia.doe.gov/cneaf/coal/page/special/feature.html (accessed August 2006). (The data for 2005 are preliminary).
16. Energy Information Administration. Energy Kid's Page. Coal—A Fossil Fuel. Available at: www.eia.doe.gov/kids/energyfacts/sources/non-renewable/coal.html (accessed April 2006).

Coal-to-Liquid Fuels

Graham Parker
Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.

Abstract
A chemical process used for turning coal into liquid fuels that has the potential for producing hundreds of thousands of barrels per day of hydrocarbon liquids and other byproducts—including electricity—is described. The key to converting coal to liquids is the Fischer–Tropsch (FT) process, which was invented in Germany in the 1920s. This process is used today in full-scale production plants in South Africa and it is being planned for use in plants in many other parts of world. A coal-to-liquids (CTL) industry is highly valued because of the security in using domestic sources of supply (coal) to produce hydrocarbons, in an environmentally acceptable process, that can be blended and refined into liquid fuels and transported to the end-user. In particular, FT fuels can play a significant role in providing a fuel currently used in the transportation industry and thus reducing dependence on imported petroleum and other refined transportation fuel products. This is of particular importance to the United States, which has an abundance of coal.

INTRODUCTION

A coal-to-liquids (CTL) plant is a chemical process plant that converts conventional pulverized coal to carbonaceous liquid fuels and byproduct hydrogen-based gases. These fuels are produced through a process that first converts the coal to coal–gas (or synthetic gas [syngas]) via conventional coal gasification and then converts liquids from the gas via the Fischer–Tropsch (FT) process. Depending on the coal quality and the way the plant is configured and operated, the CTL plant using the FT process can produce significant quantifies of light- to mid-grade, high-value hydrocarbons along with other products such as naphtha, waxes, ammonia, hydrogen, and methane. Coal-to-liquids plants are often designed to produce ∼2/3 liquid fuels and ∼1/3 chemicals such as naphtha and ammonia. One of the key products from a CTL plant (that includes post-processing or refining of FT liquid products) is high-quality/low-sulfur diesel fuel.

The critical components of a CTL plant are the coal gasifier, the enrichment of the synthetic gas to increase the hydrogen/carbon monoxide ratio (H_2/CO ratio), and the selected FT process reactor. There are many options for the critical components and component configuration of a CTL plant. In particular, there are at least eight industry-proven gasifiers, primarily used for production of only pipeline-quality natural gas, and at least three commercial production FT processes.

Keywords: Fischer–Tropsch; Coal; Petroleum; Gasification; Diesel; Hydrocarbons; Catalyst; Cetane; Carbon dioxide; Sulfur.

PRODUCING LIQUIDS FROM COAL WITH THE FISCHER–TROPSCH PROCESS

The FT process was developed in the 1920s in Germany. Inventors Franz Fischer and Hans Tropsch developed a process to convert carbon monoxide (CO) and hydrogen (H) to liquid hydrocarbons using iron (Fe) and cobalt (Co) catalysts. The temperature, pressure, and catalyst determine whether a light or heavy liquid fuel is produced. During World War II, petroleum-poor but coal-rich Germany used the FT process to supply its war machine with diesel and aviation fuel after allied forces cut off petroleum imports. Germany's yearly synthetic oil production reached more than 90 million tons in 1944.

The FT process was (and still is) used to produce most of South Africa's diesel fuel during that country's isolation under apartheid. The South African company Sasol Ltd. has produced about 1.5 billion barrels of synthetic fuel from about 800 million tons of coal since 1955 and continues to supply about 28% of that nation's fuel needs from coal.[1]

A typical CTL plant configuration using the FT process is shown in Fig. 1. The FT process is comparable with a polymerization process, resulting in a distribution of chain-lengths of the products from the process. In general, the product range includes the light hydrocarbons methane (CH_4) and ethane (C_2H_6), propane (C_3H_8), butane (C_4H_{10}), gasoline (C_5H_{12}–$C_{12}H_{26}$), diesel fuel ($C_{10}H_{22}$–$C_{15}H_{32}$), and other long-chained hydrocarbons/waxes ($>C_{15}$). The distribution of the products depends on the FT catalyst used and the process operation conditions (temperature, pressure, and residence time).[2]

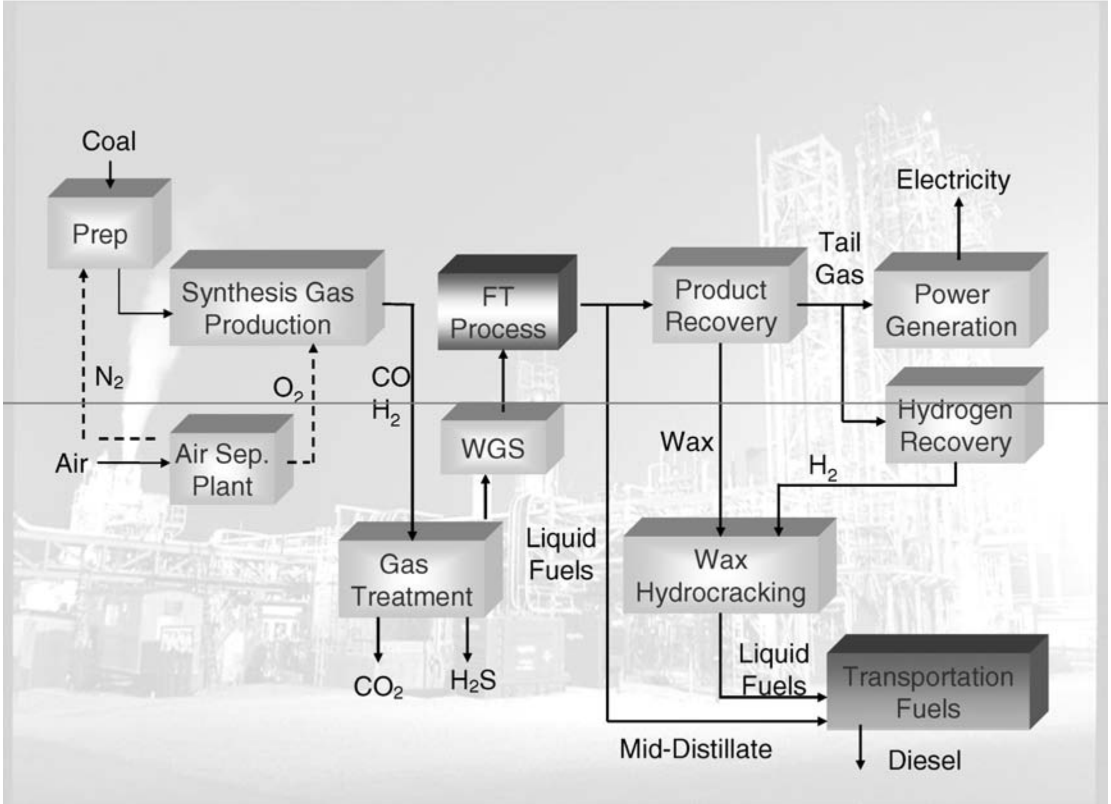


Fig. 1 Process flow diagram of a typical coal-to-liquids (CTL) plant using the Fischer–Tropsch (FT) process.

The FT process involves the use of slurry-bubble-column (slurry phase) systems using either cobalt-based (Co) or iron-based (Fe) catalysts. With these catalysts, two operating temperatures are available: low and high. For either catalyst, the FT process is exothermic. The FT reaction vessels must be actively cooled to keep them at optimal temperature for the reaction. The heat energy released by the FT process is not at a high enough temperature (200°C–300°C) to drive the production of syngas (the upstream gasification process); but it is adequate for downstream power generation if polygeneration is included in a CTL plant. (Note: polygeneration refers to the production of electricity in combustion turbines from the CTL process waste heat [primarily from the FT process] and tail gas [primarily methane] from the FT process not used to produce desired hydrocarbon products.)

Iron-based catalysts are the preferred catalysts for FT when using low CO/H_2 ratio synthesis gases derived from modern coal gasifiers. This is because in addition to reasonable FT activity, FT catalysts also possess high-water gas shift (WGS) activity. In the low temperature range, the iron catalyst can produce 50%–75% wax products. When operating at high temperatures (above 220°C), the gaseous and liquid products are highly olefinic. Another byproduct stream of the FT process is hydrogen gas. Currently, the low-temperature iron FT catalyst produces smaller quantities of products that can be used as chemicals than the high-temperature operation. A drawback with the use of Fe catalysts is their tendency to undergo attrition. This can cause fouling/plugging of downstream filters and equipment, making the separation of catalysts from the oil/wax product very difficult, if not impossible, and resulting in a steady loss of catalysts from the reactor. Iron catalysts have a higher tolerance for sulfur, are less expensive than Co catalysts, and produce more olefin products and alcohols. The lifetime of the Fe catalyst is short and in commercial installations it is generally limited to eight weeks. The Fe catalyst must then be replaced and the spent catalyst disposed of.[3]

A Co-based catalyst gives the benefit of greater management of throughput and output product selection. Co-based catalysts are often utilized because of their high FT activity, C^{+5} hydrocarbon selectivity, low WGS activity, and relatively low cost compared to Fe catalysts. The Co catalyst produces high-boiling, waxy products when operated in a low-temperature range, but attempts to operate in a high-temperature range (above about 220°C) result in the production of too much methane to be a viable option for producing liquid fuels.

Cobalt catalysts have the advantage of a higher conversion rate and a longer life (over five years); however, Co catalysts are less tolerant to sulfur and thus the upstream cleaning processes after gasification must remove most of the sulfur from the syngas. In general, the Co catalysts are more reactive for hydrogenation and therefore produce less unsaturated hydrocarbons and alcohols compared to iron catalysts. Processes have been developed to efficiently and cost effectively reactivate/regenerate and reuse a Co catalyst.

CRITICAL DRIVERS FOR A CTL INDUSTRY

There is considerable data—much of it from the oil industry—that indicates the proven reserves of crude oil and natural gas will sustain the world demand for ~100 years at current and future predicted consumption rates. There are, however, sufficient coal reserves to sustain world demand for liquid fuels for nearly 200 years. Coal is predicted to last twice as long as the combined proven crude petroleum and natural gas reserves at current usage rates. This is under a projected annual worldwide consumption of coal of 2.2 billion tons/year, or a ~1.5%/year increase.[4]

The current and future price of coal is relatively stable compared to other fossil fuels because of coal's abundance combined with the regional distribution of very large coal reserves. Currently there are about 273 billion short tons (bst) of coal reserves in North America, 173 bst in Russia and the former Soviet Union, 126 bst in China, 93 bst in India, and 90 bst in Australia.[4]

The resurgence in the interest in coal as a viable fossil fuel for direct combustion to produce electricity as well as for the production of liquid fuels is a result of the following energy dynamics:

- Steep increases and unpredictable spikes in crude oil and natural gas prices.
- A decline in domestic oil and natural gas production in those economies with high and increasing energy demands (particularly the United States, China, and India).
- Limitation in current domestic petroleum refining capacity and ability to site and build new refineries.
- The unstable political situation in the oil- and gas-rich Middle East.
- New technology developments in clean coal technologies, including coal combustion and CTL.

In addition, the production of liquid fuels from alternative and indigenous sources is addressed in the Energy Policy Act of 2005 Pub. L. 109-58, and is the linchpin of President Bush's "Advanced Energy Initiative" goal of new fuel-production technologies to replace more than 75% of our oil imports from the Middle East by 2025.

CTL PRODUCTS: DIESEL FUEL

The primary product from a CTL plant is high-quality diesel fuel. The diesel fuel from the FT process will require post-processing to make it compatible and comparable with diesel fuel derived from conventional oil refining. Typical diesel fuel characteristics from several sources are shown in Table 1.

With respect to the production of diesel fuel from a CTL plant, process conditions can be selected to produce maximum amounts of products in the diesel range. However, an even higher yield of diesel fuel

Table 1 Diesel product specifications

Property	Typical 2005 U.S. No. 2 diesel fuel	Petroleum-derived	Tar sands-derived	Fischer–Tropsch	Biodiesel
Debsity, g/cm^3	0.85	0.84	0.85	0.77	0.88
T10/T90°F	400/600	400/600	400/600	400/600	400/600
Cetane	45	48	42 (45[a])	75	56
Sulfur, ppm	400	12	12 (400 today)	3	12 (25 today)
Aromatics, %	35	32	20	0	0
Paraffins, %	35	37	20	100	0
	No. 2 diesel fuel	No. 2 diesel fuel	No. 2 diesel fuel	No. 2 diesel fuel or component	No. 2 diesel fuel or component
		Commercially available today	Commercially available today	Commercially available today	Commercially available today

[a]With Cetane improver.

can be achieved when the FT synthesis is optimized toward production of wax. Subsequently, the wax can be selectively post-processed (hydrocracking similar to processes in a refinery) to yield predominantly diesel fuel.

The resulting FT fuels from either a Fe or Co catalyst-based process are cleaner-burning than similar fuels from crude oil refining because many of the impurities are removed during the synthesis. The resulting fuels are colorless, odorless, and low in toxicity. Fischer–Tropsch fuels can be used in conventional diesel engines and have improved combustion, which reduces emissions. Fischer–Tropsch fuels have a higher cetane index and less sulfur (<5 ppm), nitrogen oxide, carbon monoxide and particulate matter emissions than petroleum fuels. In addition, the entire coal-to-liquid process is designed to remove other containments typical in fossil fuel combustion, including sulfur and carbon dioxide.

Diesel fuels from a CTL plant are being used as both a neat (a neat CTL diesel fuel is one that is introduced into the distribution system and used directly in a combustion engine without further processing or blending) fuel and as a blending fuel with conventional diesel fuel produced from petroleum processing in a refinery. The blended fuel can help refiners meet current and future sulfur standards as well as stretch the diesel fuel manufactured from conventional petroleum sources. (Note: from 2004 to 2006, the governments in some parts of Japan, Australia, and the European Union [EU] have limited sulfur in highway diesel to no more than 50 ppm. In 2006, the maximum falls to 15 ppm in some parts of the United States. Japan and the EU are expected to further restrict sulfur content to as low as 10 ppm. The cost for refiners to meet these sulfur limitations can range from $1.00 to as much as $3.00 per barrel, thus making the low sulfur diesel from a CTL process a valuable product for blending.)

The market for such fuel would include the domestic transportation industry, the department of defense (DOD), and potentially the agriculture industry. Additional post-FT processing and blending would be required and could produce other fuels, such as JP-4 and JP-8, which could be marketed to the domestic airline industry and DOD. The DOD alone requires over 3,00,000 barrels per day (bpd) of domestic diesel fuel for its operations and desires to supply this requirement with a domestic and secure resource. The Rocky Mountain (Western) states' usage will require nearly 2,00,000 bpd by 2010 and thus this is another large market. Still another large market for high-quality (low sulfur California air resources board [CARB]) diesel fuel will be California, where predictions show the state consuming over 3 billion gallons/year of diesel by 2010. (Note: current specifications for CARB diesel for use in California will be more stringent than the 2007 U.S. Environmental Protection Agency [EPA] diesel. Fischer–Tropsch diesel will be able to meet these specifications.)

Other products from a CTL plant that can be used in a combustion turbine to produce electricity and steam include elemental sulfur, naphtha, waxes, and tail gas (primarily C_4–C_6).

FT PROCESSES PROVIDERS

The Sasol Ltd. South Africa FT technology is the most mature full-scale technology in operation today, having been used in South Africa since 1955 and proposed to be used in several international CTL projects. The Sasol II and III CTL plants currently have the capacity to produce 1,50,000 bpd crude oil equivalent liquids.

There are currently three U.S.-based companies with FT technology that can be used in a CTL plant. Each company uses a proprietary technology. However, because there are no full-scale CTL plants in operation in the United States today, none of these three FT technologies have been deployed in a production-scale plant. These include:

Syntroleum Corporation
Tulsa, Oklahoma 74107
www.syntroleum.com

Syntroleum develops, owns, and licenses a proprietary process for converting natural gas to synthetic liquid hydrocarbons, known as gas-to-liquids (GTL) technology, as well as CTL technologies. For Syntroleum's GTL projects, they have executed an agreement with Exxon-Mobile to use ExxonMobile GTL patents to produce and sell fuels from natural gas, coal, or other carbonaceous substances.

Syntroleum is currently developing CTL projects with a recently announced project in Queensland, Australia. The CTL project will include removal and sequestration of CO_2 and the production of FT diesel.

Renergy Technologies Ltd. (Rentech)
Denver, Colorado 80202
www.rentech.com

Rentech began developing FT-fuel technology in 1981. It has designed FT plants in the range of 2000–40,000 bpd for potential projects in Bolivia, Indonesia, and the United States.

In June 2004, Rentech announced that it had entered into a contract with the Wyoming Business Council to perform engineering design and economic study using Rentech's patented and proprietary FT technology. The analysis evaluates the economic viability of constructing a mine-mouth plant capable of producing

10,000–12,000 bpd of ultra-low sulfur FT diesel for distribution in Wyoming, California, and other Western states. It is estimated that the facility will require about 3 million tons per year of Wyoming Powder River Basin coal for every 10,000 bpd of fuels production. The study also considered various levels of cogeneration of electric power for sale to the local transmission grid.[5]

Fischer–Tropsch projects based on coal are currently under development in Illinois, Kentucky, and Mississippi. The CTL plant in Kentucky is targeted at 57,000 bpd. In Illinois, Rentech has completed a study to convert an existing natural-gas-based ammonia plant into an integrated plant producing ammonia (900–950 tons/day), 1800–2000 bpd FT fuels, and ~10 MW exported electric power using high-sulfur Illinois coal as its feedstock. In Mississippi, Rentech has entered into an agreement with the Adams County Board of Supervisors to negotiate a contract under which Rentech would lease a site for a 10,000-bpd CTL plant producing primarily diesel fuel by 2010.

Headwaters technology innovation group (HTIG), Inc.
Lawrenceville, NJ 08648
www.htigrp.com

Headwaters technology innovation group is a wholly-owned subsidiary of Headwaters Incorporated (www.headwaters.com) that promotes and licenses technology for CTL projects using FT technology. Headwaters technology innovation group has developed an iron-based catalyst that is ideally suited for processing coal-derived syngas (synthetic gas) into ultra-clean liquid fuels. Headwaters technology innovation group has patents covering catalyst manufacturing, slurry phase reactor design and operation, production of FT liquids for fuel and chemical feedstocks, and the co-production of ammonia and FT liquids.

CURRENT AND PLANNED CTL INDUSTRY

To date, there is limited experience in full-scale (tens of thousands of bpd) production of liquid fuels from coal. Most experience in full-scale CTL plant operation is with the Sasol Ltd. operation.

A number of pilot-scale (up to hundreds of bpd) plants have been constructed and operated in the United States. These include a 35-bpd plant operating from 1975 to 1985 in Pennsylvania, a 35-bpd pilot plant operated by Air Products and Chemicals in 1984, a 230-bpd plant operated in Colorado by Rentech, a 70-bpd plant in Washington operated in 1998 by Syntroleum/ARCO, and a 2-bpd pilot plant in Oklahoma operated by Syntroleum from 1989 to 1990.

The technology challenge in a CTL plant focuses on the FT process and particularly the FT process fed with gasified U.S. coals. Experience in South Africa is with South African coal with unique coal qualities and with natural gas as a feedstock GTL. There is a dearth of experience in full-scale operation of a CTL plant in the United States with any FT process, and with the pilot/demonstration CTL plants, there has been limited testing of the qualities and performance of the fuels.

In addition to the projects noted above, there are a number of planned CTL projects under development around the world, as discussed below.

North America

A number of states are currently either considering or actively developing CTL plants using the FT process, including Kentucky, North Dakota, Mississippi, Missouri, Montana, Ohio, Pennsylvania, West Virginia, and Wyoming. Most of these projects are being developed as consortiums or partnerships of coal companies, gasification/FT suppliers, architect/engineering firms, universities, energy technology centers, and state governments.

Australia

The Australian Power and Energy Limited/Victorian Power and Liquids Project (APEL/VPLP) CTL project was planned in Australia as a joint venture between Australian Power and Energy Limited and Syntroleum. This project includes coproduction of power and hydrocarbon liquids from brown coal in the Labtrobe Valley in the state of Victoria. The initial phase of development envisions a 52,000- bpd plant with CO_2 capture and sequestration via subsurface injection.

Asia

Sasol is under discussions with China to build several CTL plants and could also take equity stakes of up to 50% in two proposed Chinese CTL projects. The Chinese proposal is part of a joint venture between Foster-Wheeler/Sasol and Huanqui for two facilities producing 80,000 bbl of liquids per day per site at the Ningxia autonomous region and the Shaanxi province, both in the coal-rich western part of China. An additional CTL/FT demonstration plant of unknown size is being planned by HTIG in Mongolia, China. There are also plans for an 80,000-bpd plant in Indonesia in partnership with Sasol.

India and Pakistan

There is considerable interest from coal companies in India and Pakistan in the Sasol CTL technology. Significant coal reserves in India and Pakistan could help to reduce dependence on imported crude oil.

SITING AND OPERATING CTL PLANTS

The development pathway for CTL plants is uncertain, given the myriad of choices of proven full-scale gasification processes that must be integrated with unproven (on a full-scale) FT processes other than the Sasol Ltd. Fischer–Tropsch process—which has only been used in a fullscale plant with South African coal. Although a CTL plant has the potential for producing tens of thousands of barrels of clean fuel (and other byproducts) in an environmentally-friendly process, there are a number of engineering, infrastructure, and institutional issues that need to be resolved before a full-scale plant is viable in the United States. These include but are not limited to improved materials/catalyst performance and reaction mechanisms for the FT process; permitting, siting, and regulating a first-of-a-kind plant that is neither a conventional coal-fired power plant nor a conventional chemical plant nor a conventional oil refinery; water use and water treatment requirements; securing a coal supply contract under high-demand conditions for coal for power plants; securing an FT-based fuel outtake contract with both a floor and ceiling price; cost-effective carbon capture and other emissions treatment strategies; and optimizing the plant output products that include liquid fuels, naphtha (feedstock for chemical production), ammonia (for fertilizer production), and electricity.

COAL FUEL SUPPLY

In the United States, there are vast deposits of coal—deposits more extensive than those of natural gas and petroleum, the other major fossil fuels. Identified resources include the demonstrated reserve base (DRB), which is comprised of coal resources that have been mapped within specified levels of reliability and accuracy and that occur in coal beds meeting minimum criteria for thickness and depth from the surface that may support economic mining under current technologies.

A typical CTL plant with the capacity of 10,000 bpd would require 10,000–15,000 tons of coal/day. Such a plant operating at 90+% capacity would consume 3.5–5 million tons of coal/year and produce over 3 million barrels of diesel fuels plus other marketable byproducts. It is anticipated that multiple plants could increase production capacity to 1,00,000 bpd by 2012, 1 million bpd by 2025, and 2–3 million bpd by 2035, and that the coal mining and coal transportation industry would be able to accommodate such production levels.

There are three major coal-producing regions in the United States: Appalachian (primarily Ohio, West Virginia, Kentucky, Tennessee, and Pennsylvania), Interior (which includes the Gulf Region and the Illinois Basin), and Western (which includes the Powder River Basin/Colorado Plateau and Northern Great Plains).[6]

U.S. coal production in 2004 totaled 1112.1 million short tons and was divided among the regions as follows:

- Appalachian: 389.9 short tons
- Western: 575.2 short tons
- Interior: 146.0 short tons

Of the total coal produced in 2004, 1.016 short tons (91%) were used to generate electricity.[4]

The actual proportion of coal resources that can be mined and recovered economically from undisturbed deposits varies from less than 40% in some underground mines to more than 90% at some surface mines. In some underground mines, much of the coal is left untouched as pillars, required to prevent surface collapse. Adverse geologic features such as folding, faulting, and interlayered rock strata limit the amount of coal that can be recovered at some underground and surface mines.

COAL MINING AND TRANSPORT

There will be a significant increase in coal mining once the CTL industry matures and production reaches estimated full-scale operation by 2035. An estimated additional 2.5–4 million tons of coal per day will be required to supply CTL plants operational by 2035. A mature CTL industry would require an additional 2%–3% production increase over 2004 levels and this can be adequately handled by the coal industry. For example, current production in the Western Region's Powder River Basin is over 350 million tons/year, thus this increased demand would add 8%–12% additional demand if all the additional demand for coal were supplied by this region.[4] This additional demand can readily be provided by the existing mines, thus no new mines would need to be permitted.

There is, however, a current limitation to the amount of coal that can be transported over existing rail lines. Transportation of coal from the mine to the consumer continues to be an issue for the industry. The majority of coal in the United States is moved by railroads exclusively or in tandem with another method of transportation.

A nearby high-speed rail line (and connecting rail spur to the plant) would be required for coal transport capable of transporting ∼100 coal cars/day for a 10,000–15,000-bpd plant. Petroleum pipelines are the preferred mode of moving FT diesel fuel; however, depending on the location of the CTL plant, rail transportation may also be the best alternative for transporting the diesel and possibly other plant byproducts to a refinery or end-user. A train with 40–45 tank cars/day would be required to transport the diesel fuel from a 10,000-bpd plant. Additional transportation modes (tanker truck, tanker car, or pipeline) would be required to haul away other products such as sulfur (truck transport), carbon for sequestration (pipeline transport), ammonia (road or rail transport), and possibly naphtha as a blending stock for gasoline refining (road or pipeline transport). Environmental issues related to increased coal transport (noise, dust) would need to be addressed by each state and community through which the trains would pass.

AIR QUALITY

A CTL plant emits far fewer criteria pollutants into the atmosphere than even the best-controlled and most efficient coal combustion power plant. Sulfur (as H_2S) and mercury (as elemental mercury and captured in impregnated activated carbon absorbent) are removed during the gasification. More than 99% of the sulfur and mercury are removed prior to producing liquids in the FT process.

A significant amount of carbon can also be captured as CO_2, with the percentage of carbon captured depending on how the plant is operated as well as the economics. Carbon can be further processed for sequestration and the sulfur is converted to elemental sulfur for dry disposal or sale. Carbon dioxide can also be used for enhanced oil recovery or for coal-bed methane extraction.

One of the key features of a CTL plant is the potential for substantial carbon capture and sequestration. This can make CTL plants environmentally preferable to combustion plants. Typical strategies for sequestering CO_2 include physical trapping, hydrodynamic trapping, solubility trapping, and mineral trapping. These sequestering and use of CO_2 strategies are illustrated in Fig. 2.

There are additional emissions of oxides of nitrogen from the combustion process used to generate electricity, chlorine, and particulates from combustion and coal transport, handling, and processing (pulverization). All

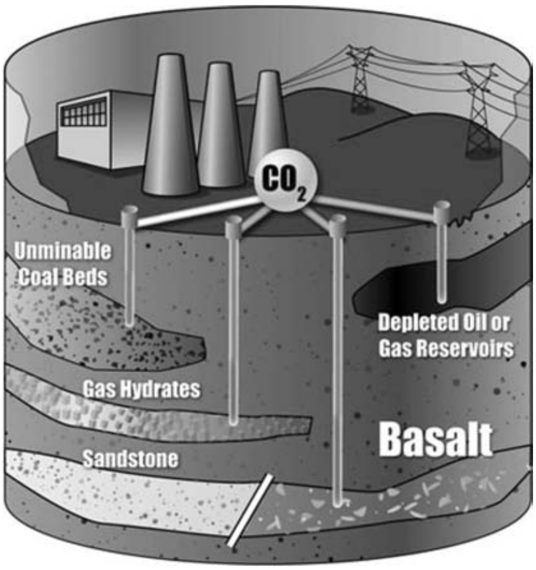


Fig. 2 Illustration of methods of sequestering and using CO_2 captured from a coal-to-liquids (CTL) plant.
Source: From Pacific Northwest National Laboratory.

of these emissions can be treated using the best available control technology (BACT). The goal of control technology for a CTL plant is to reduce the emissions to at least the level of those emitted by a conventional coal-fired power plant using the BACT to meet current air quality standards. Overall, a CTL plant allows for easier and more effective control of criteria pollutants—and additionally CO_2—compared to today's most efficient and controlled coal-fired power plants.

The only significant air quality issue in the siting and operation of a CTL plant is the potential impact the plant may have on Class I air-sheds such as those in national parks and other designated Class I areas. This would be addressed during the permitting process via air quality dispersion modeling.

WATER QUALITY

A CTL plant using high moisture content coal such as lignite (30%–50% water content) or sub-bituminous (10%–30% moisture content) will likely be a net water producer, depending on whether or not a dry or slurry feed is used for the gasifier (upstream of the FT process) and whether or not a significant amount of (excess) power is produced requiring a cooling tower. Under the plant design scenario using 30% moisture content coal, a net production of 100–200 gal/min would be discharged for a 10,000-bpd liquids plant. This water would require treatment and disposal in surface or underground wells.

Use of bituminous coal with a water content of 5%–10% or anthracite coal with a water content of $<5\%$ would likely require a water supply of 400–600 gal/min for a 10,000-bpd plant.

One achievable goal of a CTL plant is to design the plant to be "water-neutral," that is, near zero discharge and without substantial process water supply required. This will depend on the coal and specific processes used as well as the product mix (liquids and electricity generated).

LAND

The land requirement for a CTL plant is approximately 8–10 acres for a 10,000-bpd plant. This includes land for a rail spur, six days of coal storage, coal handling, the CTL plant itself, chemical and water treatment, post-FT fuel processing, electrical generation, water treatment, storage tanks, and auxiliary support equipment.

PERMITTING

The CTL plant is expected to produce discharges at the same or lower levels as the best coal-fired integrated (coal) gasification combined cycle (IGCC) power plant. Thus, the permitting process should not be onerous. Permitting would include addressing state and federal (Environmental Protection Agency) requirements for both a combustion power plant (power plant siting)—given electricity is generated on-site in a combustion turbine, and a chemical plant (industrial siting)—given the gasification and FT process are chemical processes. This includes new source performance standards and Clean Air Act requirements.

The significant solid waste to be disposed of is nontoxic, nonleachable slag (~140 kg/bbl) waste from the gasification process, and all Resource Conservation and Recovery Act (RCRA) requirements would need to be met. There are strategies that have been developed and are in use in the coal gasification industry to convert the slag into a usable and thus marketable product as well as to reuse a portion of the higher heating content slag in the gasification process. Thus, the total amount of slag can be reduced considerably if these processes are deployed.

Additional permitting would be required for (1) the construction of a rail spur; (2) coal handling/pulverization processes (fugitive dust); (3) electrical permitting for electrical supply requirements; (4) potable water supply; (5) water/wastewater discharge during construction and wastewater from the process; (6) construction and operating permits; and (7) permits related to cultural resources, native lands, National Environmental Policy Act of 1969 (NEPA), and the Federal Land and Policy Management Act.

EMPLOYMENT

A typical 10,000-bpd CTL plant would require ~1000 construction jobs with 150–200 permanent staff to operate the plant. There would be some marginal increase in employment in the coal mining and coal/diesel transportation industry, but not nearly as significant as in the plant operations. A rough estimate is an increase of 5–10 coal and transportation industry jobs for every 10,000-bpd plant.

NATURAL GAS

The CTL plant would not require significant quantities of natural gas, and this would primarily be during startup. At full production, the plant would be a net hydrocarbon producer including tail gas and naphtha from the FT process and possibly bypass gas from the gasifier. Most of the tail gas would be used to feed a combustion turbine to produce electricity for use by the plant and for the sale of any excess electricity.

ELECTRICITY

A transmission line and substation would be required to supply electricity for plant startup and an emergency generator (diesel-fueled) would be required for unanticipated plant shutdown. The CTL plant would be a net electricity generator (after use by the plant itself), using an efficient combined cycle combustion turbine generation plant ranging from 100 to 200+ MW, depending on how the plant is configured.

CONCLUSIONS

A CTL plant would be viable for countries with large reserves of relatively low-cost and readily accessible coal in grades that can be gasifiable. Stranded coal (due to location of coal quality) that cannot be easily monetized in other ways is also a viable source. The market for the liquid products should be nearby, and thus CTL plants would be most economically sited in major energy-consuming countries. In addition, there needs to be an ability to capture and sequester or otherwise use the CO_2 in order to minimize the environmental impact of the siting and operation of the plant.

Other conditions require crude oil prices above $40–$50/bbl (2005 dollars) combined with coal prices at $20–$50/ton (2005 dollars), favorable financing terms, and a short permitting process. The resulting FT fuels (diesel) must be compatible with the current diesel fuels used in engines and with a cost that is able to compete in the marketplace with diesel fuels produced from petroleum (or natural gas via a GTL process) feedstocks. If a domestic energy supply and security from using that supply are critical, there is a potential premium that could be applied to FT diesel fuel.

There remain significant challenges to a CTL industry, including the following:

- A CTL plant is highly capital-intensive and considered a risk to potential investors due to the few large-scale CTL plants operating worldwide.
- General inexperience in industry with the design, construction, and operation of a CTL plant.
- Few choices of large-scale demonstrated FT technologies and suppliers. Currently there is only one proven full-scale FT technology (Sasol) that has been used with syngas produced from coal.
- The historic volatility and uncertainty in the oil and natural gas industry makes it difficult to predict future oil and natural gas prices, and thus the economics of a CTL plant.
- The "true" cost of producing FT diesel fuel from a CTL plant and whether or not industry will pay a premium for that fuel without government guarantees or subsidies.
- The competition for coal (supply and transport) used for power generation by the utility industry potentially driving up the price of delivered coal for a CTL plant to the point of it not being economically viable.

- The dearth of industry/independent data on the long-term use of CTL/FT diesel fuel used in conventional (today's) diesel engines, particularly in the U.S. transportation industry.
- Environmental issues related to carbon capture and sequestration as well as inexperience with siting and operating a full-scale CTL plant, including issues related to "dirty coal" and the difficulty of siting of refineries.

REFERENCES

1. www.sasol.com (accessed on March 2006).
2. http://www.ecn.nl/biomassa/research/poly/ftsynthesis.en.html (accessed on March 2006).
3. www.clean-energy.us (accessed on March 2006).
4. www.eia.gov (accessed on March 2006).
5. Company mulls coal-to-diesel plant. *Casper Star Tribune*; 12 June 2004.
6. http://energy.er.usgs.gov (accessed on March 2006).

Cold Air Retrofit: Case Study☆

James P. Waltz
Energy Resource Associates, Inc., Livermore, California, U.S.A.

Abstract
The first-cost focus of new construction can often result in buildings that do not work—particularly the Heating Ventilation and Air Conditioning (HVAC) systems. This article presents an innovative cold air retrofit that corrected designed-in HVAC inadequacies in an office building. By utilizing cold air, the retrofit overcame insufficient airflow and insufficient cooling capacity at the air-handling units, thereby avoiding a massive—and very disruptive—retrofit of a fully occupied office building. The resulting retrofit cost less than half of the conventional alternatives and was easily implemented during weekend hours. It resulted in a building that actually worked for the first time since its original construction some 15 years ago. The author's more than two decades of experience in restoration and remediation of existing buildings provides some valuable insight into how to creatively and cost effectively fix nagging comfort problems in existing buildings. The author additionally provides some reflection on the energy efficiency impact of the project and energy related implications for new building designers.

INTRODUCTION

In reading this article, it is important to realize that the world of retrofit is a poorly understood and completely unique niche of the building construction industry. We find that most building owners and most design professionals don't understand this. By and large, traditional design professionals, who grew up in the world of new construction, are ill equipped to face the constraints of fixing problems in existing buildings—short of wholesale replacement of systems (which gets us right back to "clean sheet" or new construction design, right?). In the world of retrofit, all those little problems that were resolved in the field by the original builders, and all those remodels and modifications need to be identified and dealt with by the retrofit engineer. Frequently, whole building testing is necessary to identify the cause of the complaints. In addition, retrofit frequently requires that the building be modified while the building is fully occupied, meaning that working conditions are difficult, and major disruption to the occupants cannot be allowed. We frequently liken it to performing a heart transplant on a marathon runner—during a marathon. The project described herein is just such a project.

THE SITUATION

We were called into the project by the service contractor of the building, who was trying to figure out how to help the owners to keep their tenants happy. They weren't very happy, as the building was uncomfortably warm nearly all year long in its northern California climate. Only during the coldest months of the year was the building comfortable. Just as in our expert testimony work, we set about to ferret out the source of the problem. What we learned was that the original designer made some fundamental conceptual errors in determining the operating parameters of the air handling equipment, which effectively resulted in undersizing of both the airflow and the cooling coils. This was in spite of the fact that he had done a good job of estimating the cooling needs of the building. The problem wasn't in the capacity of the chiller, or in the apparent capacity of the air-handling units. There was a problem, though, in the performance of the air-handling units.

There are a couple of aspects of load calculations that, as the carton character Dilbert says about nuclear energy, "can be used for good or evil." Those aspects have to do with space loads versus system loads. Astute HVAC system designers are very careful when considering these loads, realizing that any cooling load that can be kept out of the occupied space allows the designer to reduce the supply air quantity needed to cool the space, and in turn allows the use of a smaller air-handling unit. Seems pretty obvious, right?

Well, in this case the designer assumed that 100% of the heat from the lights would go into the return air instead of the space. After all, he was assuming that return-air-troffer lighting fixtures would be used, and therefore all the heat from the lights would go into the return air passing through the fixtures. On the surface this seems plausible. However, certain fixture manufacturers actually document the percent of the total heat from a fluorescent fixture that is transferred into the air stream. The highest value we've

☆ This entry originally appeared as "Making Buildings Work: A Case Study of an Innovative Cold Air Retrofit" in *Energy Engineering*, Vol. 100, No. 6, 2003. Reprinted with permission from AEE/Fairmont Press.

Keywords: Cooling capacity; Retrofits; Case studies; Air flow; Redesign.

seen so far is about 30%, and we think even that is a bit optimistic. When you consider that the lighting system heat gain can contribute as much as 40%–60% of the total heat gain in an occupied space, this had a dramatic effect on the calculated supply air cfm. Add to this the fact that the HVAC system was designed for a "shell" building and the eventual tenant build-out did not employ return air troffer lighting fixtures, you can start to get an idea of how much trouble this building was in. But there's more.

The "more" is the other effect of the designer's assumption about the space loads—its effect on the load system experiences. You see if you assume that 100% of the heat from the lights goes into the return air, instead of a return air temperature of, say 76°C, you will calculate a return air temperature of more like 86°C. This means that when combined with a fairly high ambient design temperature, the mixed air temperature will calculate out to about 88°C, instead of a more correct value like 78°C. This only becomes a problem when selecting a cooling coil for your (already undersized!) air-handling unit. Since there will be more heat transferred from 88°C air to 45°C chilled water than from 78°C air (total temperature difference is now thought to be 88–45 or 43°C rather than 78–45 or 33°C), it will appear that you can get all the cooling done that you need with a pretty small coil (i.e., fewer rows and/or fewer fins per inch). Indeed, such an undersized coil was selected by the system designer

The net–net of all the above is that by making one "fatally" wrong assumption, the system designer put into the building an air-side system that could never cool the building—and indeed it didn't.

SOLVING THE "PUZZLE"

Once we understood the root of the problem, we had to face the question of what to do about it.

The immediately obvious solution was to yank out the air handling units and replace them—what we would call the "traditional" approach. After all, this would correct the fundamental error that was made in the first place. The problems with this approach were, not surprisingly, many fold, and included:

- The air handling units were located in interior mechanical rooms in the "core" area of each floor, and replacing them would require knocking out walls and seriously disrupting the occupants and operations of the building.
- Increasing the horsepower of the air handling units would require significant cost for electrical work as all the air handling units were fed electrical power from the basement and the entire conduit and conductor riser would need to be replaced, as it had no excess capacity.
- The mechanical rooms were very cramped and there really was no room at all in them for larger air handling units. This would require re-configuring the floor plan layout of the "core," another very expensive proposition (and likely not really feasible).

Recognizing that a more traditional approach really did not constitute a suitable solution for this problem (and was likely the reason the problem had gone unresolved for 15 years), Energy Resonance Associates (ERA) set about to "re-engineer" the Heating, Ventilation and Air Conditioning (HVAC) system from the inside out, assuming that the air handling units themselves could not be replaced, nor could their fan horsepower be increased (due to the limitations of the building's power distribution system). Grinding away with a computerized coil selection program and rethinking other parts of the HVAC system, we determined that the air handling units could be made to perform by:

- Replacing the existing 4-row chilled water coils with 8-row coils of equal air pressure drop (examination of factory certified dimension drawings confirmed that they would fit in the air handling units).
- Increasing the chilled water flow through the coils (feasible with a much higher horsepower pump, and within the allowable flow rate for the chiller—and requiring more than twice the original horsepower to achieve a 30% increase in flow).
- Reducing the chilled water supply temperature (from 45°C to 40°C, also with the allowable operating parameters for the chiller).
- Installing new air handling unit temperature controls (to reset the planned very-low supply air temperature upwards during cool weather, else "cold" complaints would replace the prior "hot" complaints).

Without negatively impacting the air handling systems' air supply rate, the new system would be capable of supplying 46°C- air, thereby produce the actual cooling needed to satisfy the occupied space. As shown in the table below, some pretty interesting results can be achieved by optimizing the coil selection in particular!

Description	ROWS/FPI	EDB/EWB	LDB/LWB	EWT/LWT	GPM	APD	MBH
Existing	4/12	74/60	53.5/51.8	45/58.5	30	0.69	203
Retrofit	8/9	74/57	45.8/44.6	40/54.4	39	0.70	281

Coal–Comm

MAKING THE FIX

Upon completion of the study, ERA was engaged to prepare final installation documents. This work was performed in collaboration with the owner's selected contractor so as to achieve maximum integration of design concepts and the contractor's working knowledge of the building (the contractor had the service contract for the building). Final selection of equipment was made, simplified installation drawings were prepared, and the project installed and put into operation over a 90-day period, including startup. No tenant disruption was caused during the installation (which would have been the case had the conventional approach of replacing the air handling units been followed). Upon completion of the project, the building's HVAC systems provided comfort for the first time in the 15-year life of the building! The utterly prosaic business of HVAC engineering doesn't get any more exciting than this.

SOME INTERESTING CONCLUSIONS

One of the lessons that can be learned from this project is that the age-old tradition of linking engineering fees to construction cost—our traditional way of paying design professionals—would not have allowed this project to take place. After all, it took a lot of engineering to avoid spending money. So engineering fees went up, and construction costs went down, making the engineer's fees look "large" as a percent of construction costs. Many building owners would insist that less money be spent on engineering—with the result that the engineer is forced to get his eraser out to create a "clean sheet of paper" and do a very simple design, that doubles or triples construction costs. Voila! The engineer's fees look "small" as a percent of construction. Building owners, take heed.

Another, perhaps more technical lesson to be learned is that by understanding the essential nature of the engineering problem being faced, it is often possible to re-engineer a system from the inside out and make it work, even when it seems impossible. Design engineers, take heed.

A final lesson for new building HVAC designers is that if you want to build a little "safety" into your HVAC system, selecting a cooling coil with more rows (and perhaps a few less fins per inch) is really, really cheap "insurance."

Readers may have noticed that this retrofit would likely have the effect of increasing the energy use of this building. Our charter from the building owner on this project was to make the building work. They were not at all interested in energy conservation—much to the contrary. The truth is, even in today's energy sensitive environment, making buildings work, i.e., having them provide the function they were intended to provide (a comfortable and productive work environment), is equally, if not more important, than saving a few dollars on the utility bill (and this coming from an award-winning energy engineer).

For energy engineers, the lesson is that cold air works—and offers some interesting energy saving opportunities. If you were designing this building from scratch, using a conventional design would have required a larger (and probably more powerful) air handling unit, so cold air would have saved a lot of air circulation energy. Since the fans run whenever the building is occupied (or even longer), these savings would be dramatic. While we ran the chiller at a colder evaporator temperature, in a new building design the cooling tower could have been oversized (at relatively minimal cost) to compensate and keep the total chiller "lift" (which is what you pay for in terms of chiller power) the same as a conventional design, or even better. In this retrofit we had to increase pump power—rather dramatically. In a new design, a nominal (and relatively cheap) oversizing of the piping could have been done, and the system configured for variable flow (see our contracting business article on our web site for this) and the pumping power kept to a minimum as well. Finally, an energy engineer "worth his salt" would include variable speed drives on the fans and a digital control system to precisely reset supply air temperature (to minimize reheating) and manage the operation of the chilled water system and optimize run hours of the entire HVAC system.

Combined Heat and Power (CHP): Integration with Industrial Processes

James A. Clark
Clark Energy, Inc., Broomall, Pennsylvania, U.S.A.

Abstract
This entry discusses the integration of combined heat and power (CHP), otherwise known as cogeneration, with industrial processes. It builds on other entries in this encyclopedia that discuss the basics of CHP or cogeneration.

INTRODUCTION

This section discusses the integration of CHP, also referred to as cogeneration, with industrial processes. Topics discussed in this section include:

- Cogeneration unit location
- Industrial electric power systems
- Generator voltage selection
- Synchronous vs. induction generators
- Industrial thermal considerations
- Industrial process examples utilizing unique solutions
- Exhaust gas condensation solutions
- Cogeneration and interruptible fuel rates
- Future considerations

COGENERATION UNIT LOCATION

The location of the cogeneration unit is usually a function of the location of the following:

- Main electric switchgear
- Large electric branch circuit switchgear
- Natural gas lines or oil storage tank
- Thermal processes
- Exhaust stack
- Ancillary services, such as power, water, and sewer
- Utility relay and control location

A location that minimizes the cost of electric lines, piping, and other components required to supply the above services is the optimal solution. Industrial processes and branch circuit locations differentiate the location decision from nonindustrial sites. The cost of these components can have a substantial impact on the project cost.

Keywords: CHP; Cogeneration; Industrial; Processes; Heat recovery; Energy recovery; Interruptible gas.

INDUSTRIAL ELECTRIC POWER SYSTEMS

Industrial electric power systems typically involve multiple levels of voltage reduction through multiple levels of electric transformers. Most of these transformers are owned by the customer. By contrast, small commercial and residential customers are typically provided with one voltage supplied by a utility-owned transformer.

GENERATOR VOLTAGE

The generator voltage depends on the voltage of the industrial customer and the generator size. Manufacturers supply cogeneration units in multiple voltage sizes; however, the size range is generally limited to practical voltages. Because the size of the generator is inversely proportional to the voltage, there is a limitation to the minimum voltage based on the size and cost of the resultant generator.

Generator voltage should generally be matched to either the voltage after the main transformer or the voltage of the branch circuit, where most of the generator output will be utilized. In this manner, the electricity produced will travel through the least amount of transformers. Each time electricity travels through an electric transformer, approximately 2% or more of the electricity is given to transformer losses. When calculating the benefits of electricity production, these losses (or avoidance of these losses) should be considered. In nonindustrial applications, the cogeneration unit is typically installed at the electric service entrance. In industrial applications, the cogeneration unit can be installed in one of the branch circuits.

SYNCHRONOUS VS. INDUCTION GENERATORS

Generators are essentially electric motors in reverse; rather than using electricity, a generator produces electricity. The technology behind both can easily be described as an electric magnet spinning in a casing surrounded by wires.

A prime difference between synchronous vs. induction generators is that the synchronous generator is self-excited. This means that the power for the magnet is supplied by the generator and its control system. By contrast, an induction generator is excited by the utility. When utility power is lost, the loss of excitation power will theoretically shut down the induction generator. However, there is a chance that other electric system components could provide the excitation required to start the generator. Therefore, in addition to the electric utility engineers, a competent professional engineer should approve the design of the cogeneration installation.

Cogeneration systems require protective relays on the power system that will shut down the cogeneration installation in the event of a power interruption. They are not designed to run in power outages. They must be shut down or utility personnel will be endangered when they try to restore power to a nonpowered electric line. Many utilities have less stringent protective relay requirements for induction systems for the reasons explained above. The extra costs, however, can be prohibitive. It is important for owners to determine the costs of protective relays and other interconnection requirements before commencing construction. Many owners have been surprised after the fact and had to spend additional money to correct deficiencies.

Power factor is another consideratcion in the selection of induction vs. synchronous generators. Because induction generators use utility reactive power to excite the generator, the site power factor can degrade significantly. By using more reactive power while simultaneously reducing real (kW) power, there is a double effect on power factor. Why should an owner be concerned about power factor? Because many utilities measure and charge industrial users for reactive power and/or low power factor.

INDUSTRIAL THERMAL CONSIDERATIONS

Most cogeneration installations depend on almost full utilization of the thermal output of the plant. In order to utilize the full thermal output, the thermal usages and processes must be considered carefully. The thermal output each and every hour, as well as the temperature of the thermal output as it compares to the temperature of the industrial processes, must be examined.

Industrial thermal considerations include:

- Steam vs. hot water
- The temperature of cogeneration thermal output and industrial processes

Steam vs. Hot Water

The thermal output of gas turbine installations is usually steam, whereas the output of reciprocating engines is generally hot water in the 200°F range. There are exceptions to this general rule. For industrial processes that require higher temperatures for thermal processes, a gas turbine selection may be necessary. As described below, care should be taken to ensure that the thermal output can be utilized. In other words, if 240°F thermal energy is needed, a reciprocating engine may not be able to meet the requirements unless properly designed.

Temperature of Cogeneration Thermal Output and Industrial Processes

As mentioned above, the temperature of the thermal output of the cogeneration system may not be utilized in all thermal industrial processes. Simply put, the cogeneration thermal output should be at a higher temperature that the industrial process temperature. Otherwise, the owner may realize less savings than anticipated.

The engineer must perform a thermal balance to ensure that the thermal output of the cogeneration system is less than the industrial process requirements for all hours. The author likes to ask what the thermal needs are in July at midnight. At that time, comfort heating requirements are nonexistent and many industrial thermal processes may be shut down.

A common error occurs when thermal balance is made on an annual, rather than an hourly, basis. Even if the annual thermal output of the cogeneration system matches the annual thermal input of the industrial thermal processes, there are usually hours when excess thermal output must be discarded.

Thermal storage is a potential solution for the hourly ups and downs of thermal energy requirements. The author has found, however, that this solution often falls short; if more than an hour or two of storage is needed, the amount of storage required is both too costly and impractical.

INDUSTRIAL PROCESS EXAMPLES UTILIZING UNIQUE SOLUTIONS

The author has been involved in many installations where finding a unique thermal application has rescued a financially challenged project. A few of these applications are explained below.

Paper Plant

In a paper plant, three thermal industrial processes were added to existing systems and a steam generation system was designed and added to the reciprocating exhaust gas stream. In general, a better solution is to reduce the cogeneration system size by 10%–20%, rather than to add a complicated steam generation system. A hot water recovery system on the exhaust gas stream is generally more cost effective. However, in this installation, a steam generation system was added. Exhaust gas was fed into a steam generator to produce steam directly at 100 psig. A secondary heat exchanger was added after the steam generators in order to recover additional energy.

The second unique application was a heat exchanger being added to a water pulping tank. Water and paper pulp are added in one of the first steps of the paper-making process. This water/pulp slurry was heated by direct injected steam originally. Water at 200°F was introduced to the tank via a piping serpentine installed in the tank. The water in the tank was heated to approximately 140°F. The large temperature difference allowed for the use of nonfinned piping of a reasonable length. It is very important to note that an energy balance needed to be considered. The steam used in this tank was low pressure steam that had been recovered by a flash steam energy recovery tank. Essentially, this steam was free, and if it could not be utilized elsewhere, the value of the cogeneration hot water would be zero. Because the steam was utilized elsewhere, the energy was useful.

The third unique thermal energy recovery system added to this paper plant was a hot air blower system. In the paper-making process, the water/paper slurry is eventually dried on a steam drum by being drawn over the drum by a continuous paper sheet. Steam in the drum dries off the water in the slurry, leaving only paper. By blowing hot dry air over the paper sheet, the paper dries faster. The first benefit is that less steam is needed because the hot air dries the paper. An even more important benefit is that the speed of the paper sheet in the process can be increased and more paper is produced with the same overhead and labor costs. This financial benefit can dwarf the cogeneration system savings. The physical limitations of the existing system were overcome by the addition of this production enhancing system.

Car Parts Plant

A solution considered in a car parts plant was to direct the engine exhaust directly in a parts drying process. The 1200°F exhaust gas temperature was to be directed into a large process heater, where a 300°F temperature was maintained. One ironic outcome of this solution was that project economics were actually impaired because condensation of the exhaust gas could not be implemented. The benefits of exhaust gas condensation will be described below.

Boiler Plant Air and Water Reheat

Opportunities exist at the central boiler plant at many industrial plants. The boiler water is preheated in a makeup tank and/or condensate tank. Both returning condensate water at 180°F (or less) and makeup water at 60°F represent opportunities to use cogeneration thermal output. At both of these temperatures, an added opportunity to condense the exhaust gas, as explained later, also exists.

A less practical but potential opportunity exists to preheat the combustion air to the boiler. Each 40°F increase in the boiler air temperature equates to approximately a 1% drop in boiler fuel usage. The boiler air-fuel ratio and the combustion air fan speed must often be reset. This solution is often impractical, costly, and it provides marginal returns.

Fuel Switch on an Heating, Ventilating, and Air Conditioning System

A unique and profitable solution can be to utilize excess cogeneration thermal output for a new usage. At one facility, the author's company replaced an electric resistance coil in a heating system with a hot water coil. The electric rates were approximately four times the gas rates. In effect, the thermal output had four times the value and effectively boosted the electrical efficiency of the plant. The economic benefit of the displaced electric energy was almost equal to the fueling costs of the cogeneration plant, even though it only amounted to about a quarter of the thermal output of the cogeneration plant.

EXHAUST GAS CONDENSATION SOLUTIONS

The energy that is contained in the input fuel is generally not fully recovered. The more energy that has been recovered from the exhaust gas stream, the lower the temperature of the exiting exhaust gas stream. Approximately 10% of the entire energy that is contained in the fuel can only be recovered if the water in the exhaust is condensed. The same amount of energy that it takes to boil water is available if the reverse process of condensation is conducted. Converting the water vapor in the exhaust gas stream into water releases 10% additional energy. To condense the water in the exhaust, water at a temperature less than 200°F is generally needed.

If exhaust gas is condensed, plastic or stainless steel exhaust stack materials must be used. Care must be taken in design because an exhaust temperature that is too high can melt plastic exhaust stacks. Regular steel cannot be used because sulfur in the fuel and nitrogen in the air can cause the production of sulfuric and nitric acid in the condensate. Both of these compounds are extremely corrosive to normal steel.

COGENERATION AND INTERRUPTIBLE FUEL RATES

In the 1980s and 1990s, when natural gas prices at the well head were low, cogeneration was viewed as a natural way to increase gas companies' market share. A cogeneration system with a 40% thermal efficiency would result in twice as much, or more, gas usage as an 80% efficient boiler. Further discounts were offered for interruptible fuel rates, wherein a customer agrees to reduce their gas usage upon notification from the gas company. Gas companies were able to offer these interruptible discounts because they were using spare distribution pipe capacity during nonwinter months. In the late 1990s, natural gas fuel prices began to abruptly increase. Many discount gas rates have disappeared since then.

FUTURE CONSIDERATIONS

In the future, gas rates may decrease, making cogeneration more cost effective. Cogeneration gas rates and interruptible gas rates may return, even if for only a few months each year.

Increases in the efficiencies of reciprocating engines, gas turbines, and microturbines are already taking place and should continue. The cost of technologies, such as fuel cells and alternative energy technologies, are also decreasing. Some of these technologies may become cost competitive in the near future. Fuel cells already have higher electrical efficiencies than most present cogeneration technologies, but the cost is presently much higher than turbine and reciprocating engine technology.

CONCLUSION

The integration of CHP or cogeneration into industrial processes offers unique opportunities to optimize revenue and energy savings. The installation of new industrial thermal processes can make a financially challenged cogeneration project feasible. Condensation of the exhaust gas stream can both produce more revenue and increase overall system efficiency by more than 10%. Future advances in technologies, such as fuel cells, may offer additional options.

BIBLIOGRAPHY

1. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *The Guide to Energy Management (Chapter 14)*, 5th Ed.; Fairmont Press: Lilburn, GA, 2006.
2. Turner, W.C., Ed. *Energy Management Handbook (Chapter 7),* 6th Ed., Fairmont Press: Lilburn, GA, 2006.
3. Petchers, N. *Handbook Combined Heating, cooling and Power*; Fairmont Press: Lilburn, GA, 2002.
4. Limaye, D.R., Ed. *Industrial Cogeneration Applications;* Fairmont Press: Lilburn, GA, 1987.
5. Payne, F.W., Ed. *Cogeneration Management Reference Guide;* Fairmont Press: Lilburn, GA, 1997.

Commissioning: Existing Buildings

David E. Claridge
Department of Mechanical Engineering, Associate Director, Energy Systems Laboratory, Texas A & M University, College Station, Texas, U.S.A.

Mingsheng Liu
Architectural Engineering Program, Director, Energy Systems Laboratory, Peter Kiewit Institute, University of Nebraska—Lincoln, Omaha, Nebraska, U.S.A.

W. D. Turner
Director, Energy Systems Laboratory, Texas A & M University, College Station, Texas, U.S.A.

Abstract
Commissioning an existing building is referred to by various terms, including recommissioning, retrocommissioning, and continuous commissioning® (CC®). A comprehensive study of 182 existing buildings totaling over 22,000,000 ft^2 in floor area reported average energy savings of 18% at an average cost of \$0. 41/ft^2 after they were commissioned, producing an average simple payback of 2.1 years. The commissioning process for an existing building involves steps that should include building screening, a commissioning assessment to estimate savings potential and cost, plan development and team formation, development of performance baselines, detailed measurements and commissioning measure development, implementation, and follow-up to maintain persistence. Existing building commissioning has been successfully used in energy management programs as a standalone measure, as a follow-up to the retrofit process, as a rapid payback Energy Conservation Measure (ECM) in a retrofit program, and as a means to ensure that a building meets or exceeds its energy performance goals. Very often, it is the most cost-effective single energy management option available in a large building.

INTRODUCTION

Commissioning an existing building has been shown to be a key energy management activity over the last decade, often resulting in energy savings of 10, 20 or sometimes 30% without significant capital investment. It generally provides an energy payback of less than three years. In addition, building comfort is improved, systems operate better, and maintenance cost is reduced. Commissioning measures typically require no capital investment, though the process often identifies maintenance that is required before the commissioning can be completed. Potential capital upgrades or retrofits are often identified during the commissioning activities, and knowledge gained during the process permits more accurate quantification of benefits than is possible with a typical audit. Involvement of facilities personnel in the process can also lead to improved staff technical skills.

This entry is intended to provide the reader with an overview of the costs, benefits, and process of commissioning an existing building. There is no single definition of commissioning for an existing building, so several widely used commissioning definitions are given. A short case study illustrates the changes made when an existing building is commissioned, along with its impact. This is followed by a short summary of published information on the range of costs and benefits. The major portion of the article describes the commissioning process used by the authors in existing buildings so the reader can determine whether and how to implement a commissioning program. Monitoring and verification (M&V) may be very important to a successful commissioning program. Some commissioning-specific M&V issues are discussed, particularly the role of M&V in identifying the need for follow-up commissioning activities.

Keywords: Commissioning; Retrocommissioning; Recommissioning; Continuous commissioning®; Commissioning existing buildings; Monitoring and verification; Energy conservation measure.

COMMISSIONING DEFINITIONS

The commissioning of a navy ship is the order or process that makes it completely ready for active duty. Over the last two decades, the term has come to refer to the process that makes a building or some of its systems completely ready for use. In the case of existing buildings, it generally refers to a restoration or improvement in the operation or function of the building systems. A widely used short definition of new building commissioning is the process of ensuring systems are designed, installed, functionally

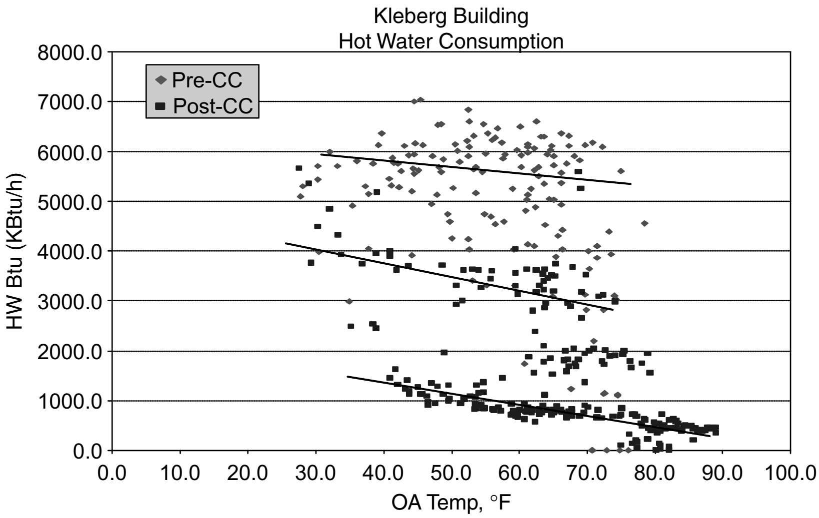


Fig. 1 Pre-CC and post-CC heating water consumption at the Kleberg building vs daily average outdoor temperature.
Source: From Proceedings of the 25th WEEC, Atlanta, GA, October 9–11 (see Ref. 4).

tested, and operated in conformance with the design intent. Commissioning begins with planning and includes design, construction, start-up, acceptance, and training and can be applied throughout the life of the building. Furthermore, the commissioning process encompasses and coordinates the traditionally separate functions of systems documentation, equipment start-up, control system calibration, testing and balancing, and performance testing.[1]

Recommissioning

Recommissioning refers to commissioning a building that has already been commissioned at least once. After a building has been commissioned during the construction process, recommissioning ensures that the building continues to operate effectively and efficiently. Buildings, even if perfectly commissioned, will normally drift away from optimum performance over time, due to system degradation, usage changes, or failure to correctly diagnose the root cause of comfort complaints. Therefore, recommissioning normally reapplies the original commissioning procedures in order to keep the building operating according to design intent, or it may modify them for current operating needs.

Optimally, recommissioning becomes part of a facility's continuing operations and maintenance (O&M) program. There is not a consensus on recommissioning frequency, but some consider that it should occur every 3–5 years. If there are frequent build-outs or changes in building use, recommissioning may need to be repeated more often.[2]

Retrocommissioning

Retrocommissioning is the first-time commissioning of an existing building. Many of the steps in the retrocommissioning process are similar to those for commissioning. Retrocommissioning, however, occurs after construction, as an independent process, and its focus is usually on energy-using equipment such as mechanical equipment and related controls. Retrocommissioning may or may not bring the building back to its original design intent, since the usage may have changed or the original design documentation may no longer exist.[2]

Continuous Commissioning

Continuous Commissioning (CC®)[3] is an ongoing process to resolve operating problems, improve comfort, optimize energy use, and identify retrofits for existing commercial and institutional buildings and central plant facilities. Continuous commissioning focuses on improving overall system control and operations for the building as it is currently utilized, and on meeting existing facility needs. Continuous commissioning is much more than an O&M program. It is not intended to ensure that a building's systems function as originally designed, but it ensures that the building and its systems operate optimally to meet the current uses of the building. As part of the CC process, a comprehensive engineering evaluation is conducted for both building functionality and system functions. Optimal operational parameters and schedules are developed based on actual building conditions and current occupancy requirements.

COMMISSIONING CASE STUDY—KLEBERG BUILDING

The Kleberg Building is a teaching/research facility on the Texas A&M campus consisting of classrooms, offices, and laboratories, with a total floor area of approximately 165,030 ft^2. A CC investigation was initiated in the summer of 1996 due to the extremely high level of simultaneous heating and cooling observed in the building.[4] Figs. 1 and 2 show daily heating and cooling

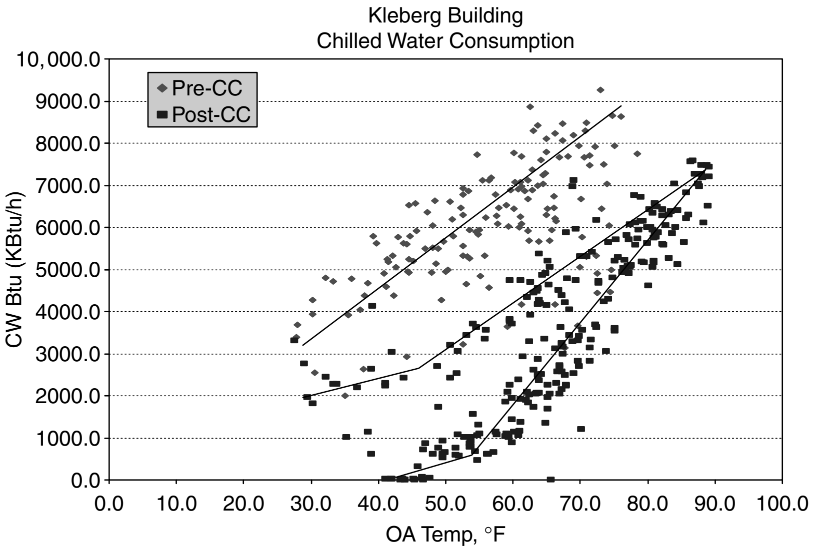


Fig. 2 Pre-CC and post-CC chilled water consumption at the Kleberg building vs daily average outdoor temperature.
Source: From Proceedings of the 25th WEEC, Atlanta, GA, October 9–11 (see Ref. 4).

consumption (expressed in average kBtu/h) as functions of daily average temperature. The pre-CC heating consumption data given in Fig. 1 show very little temperature dependence as indicated by the regression line derived from the data. Data values were typically between 5 and 6 MMBtu/h with occasional lower values. The cooling consumption is even higher (Fig. 2), though it shows more temperature dependence.

It was soon found that the preheat was operating continuously, heating the mixed air entering the cooling coil to approximately 105°F. The preheat was turned off, and heating and cooling consumption both dropped by about 2 MMBtu/h as shown by the middle clouds of data in Figs. 1 and 2. Subsequently, the building was thoroughly examined, and a comprehensive list of commissioning measures was developed and implemented. The principal measures implemented that led to reduced heating and cooling consumption were as follows:

- "Preheat to 105°F" was changed to "Preheat to 40°F."
- The cold deck schedule was changed from "55°F fixed" to "Vary from 62 to 57°F as ambient temperature varies from 40 to 60°F."
- The economizer was set to maintain mixed air at 57°F whenever the outside air was below 60°F.
- Static pressure control was reduced from 1.5 inH2O to 1.0 inH2O, and a nighttime set-back to 0.5 inH2O was implemented.
- A number of broken variable air volume terminal (VFD) boxes were replaced or repaired.
- Chilled water pump variable frequency drives (VFDs) were turned on.

These changes further reduced chilled water and heating hot water use as shown in Figs. 1 and 2 for a total annualized reduction of 63% in chilled-water use and 84% in hot-water use.

COSTS AND BENEFITS OF COMMISSIONING EXISTING BUILDINGS

The most comprehensive study of the costs and benefits of commissioning existing buildings was conducted by Mills et al.[5,6] This study examined the impact of commissioning 182 existing buildings with over 22,000,000 ft^2. The commissioning cost of these projects ranged from below \$0.10/ft^2–\$3.86/ft^2, but most were less than \$0.50/ft^2 with an average cost of \$0.41/ft^2. Savings ranged from essentially zero to 54% of total energy use, with an average of 18%. This range reflects not only differences among buildings in the potential for commissioning savings, but doubtless also includes differences in the level of commissioning applied and the skill of the commissioning providers. Simple payback times ranged from less than a month to over 20 years, with an average of 2.1 years. Fig. 3 illustrates the average payback as a function of building type and the precommissioning energy cost intensity. The sample sizes for office buildings and higher education are large enough that these averages for payback and energy savings may be representative, but the other sample sizes are so small that they may be significantly skewed by building specific and/or other factors.

Mills et al. concluded, "We find that commissioning is one of the most cost-effective means of improving energy efficiency in commercial buildings. While not a panacea, it can play a major and strategically important role in achieving national energy savings goals—with cost-effective savings potential of \$18 billion per year or more in commercial buildings across the United States."

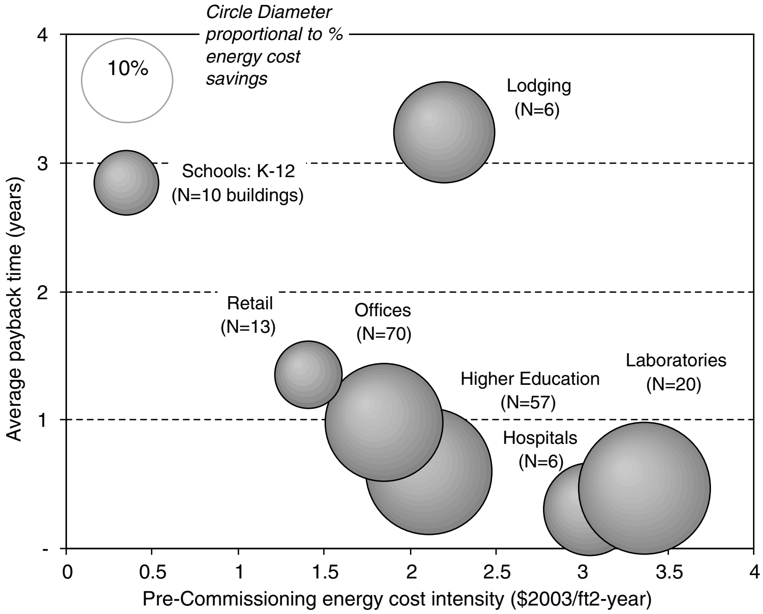


Fig. 3 Average simple payback time and percent energy savings from commissioning of existing buildings by building type.
Source: From Lawrence Berkeley National Laboratory Report No. 56637 (see Ref. 5).

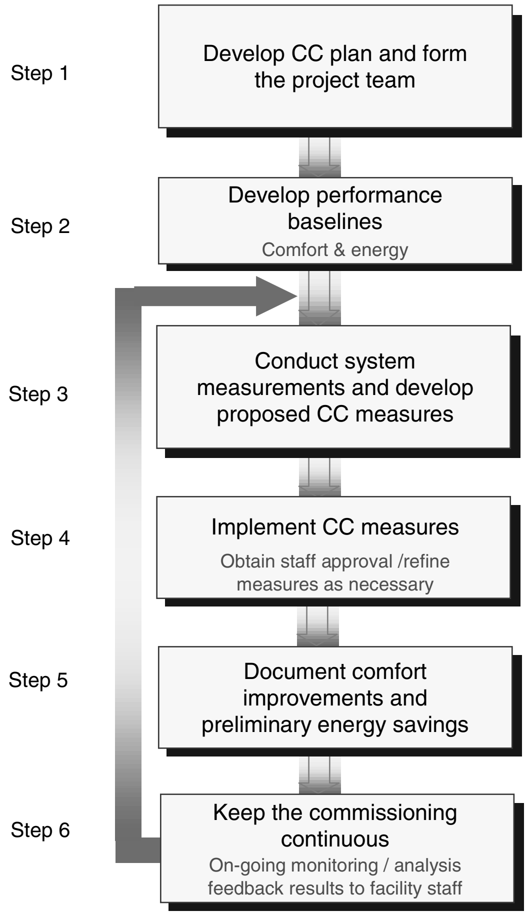


Fig. 4 Outline of phase II of the CC process: implementation and verification.
Source: From Energy Systems Laboratory.

THE COMMISSIONING PROCESS IN EXISTING BUILDINGS

There are multiple terms that describe the commissioning process for existing buildings, as noted in the previous section. Likewise, there are many adaptations of the process itself. The same practitioner will implement the process differently in different buildings, based on the budget and the owner requirements. The process described here is the process used by the chapter authors when the owner wants a thorough commissioning job. The terminology used will refer to the CC process, but many of the steps are the same for retrocommissioning or recommissioning. The model described assumes that a commissioning provider is involved, since that is normally the case. Some (or all) of the steps may be implemented by the facility staff if they have the expertise and adequate staffing levels to take on the work.

Continuous commissioning focuses on improving overall system control and operations for the building as it is currently utilized, and on meeting existing facility needs. It does not ensure that the systems function as originally designed, but ensures that the building and systems operate optimally to meet the current requirements. During the CC process, a comprehensive engineering evaluation is conducted for both building functionality and system functions. The optimal operational parameters and schedules are developed based on actual building conditions and current occupancy requirements. An integrated approach is used to implement these optimal schedules to ensure practical local and global system optimization and persistence of the improved operation schedules.

Commissioning Team

The CC team consists of a project manager, one or more CC engineers and CC technicians, and one or more designated members of the facility operating team. The primary responsibilities of the team members are shown in Table 1. The project manager can be an owner representative or a CC provider representative. It is essential that the engineers have the qualifications and experience to perform the work specified in the table. The designated facility team members generally include at least one lead heating, ventilating and air conditioning (HVAC) technician and an energy management control system (EMCS) operator or engineer. It is essential that the designated members of the facility operating team actively participate in the process and be convinced of the value of the measures proposed and implemented, or operation will rapidly revert to old practices.

Continuous Commissioning Process

The CC process consists of two phases. The first phase is the project development phase that identifies the buildings to be included in the project and develops the project scope. At the end of this phase, the CC scope is clearly defined and a CC contract is signed, as described in "Phase 1: Project Development." The second phase implements CC and verifies project performance through the six steps outlined in Fig. 4 and described in "Phase 2: CC Implementation and Verification."

Phase 1: Project Development

Step 1: *Identify Candidate Buildings*. Buildings are screened to identify those that will receive a CC assessment. Buildings that provide poor thermal comfort, consume excessive energy, or have design features of the HVAC systems that are not fully used are typically good candidates for a CC assessment. Continuous commissioning can be effectively implemented in buildings that have received energy efficiency retrofits, in newer buildings, and in existing buildings that have not received energy efficiency upgrades. In other words, virtually any building can be a potential CC candidate. The CC provider should perform a preliminary analysis to check the feasibility of using the CC process on candidate facilities before performing a CC assessment.

The following information is needed for the preliminary assessment:

- Monthly utility bills for at least 12 months.
- General building information—size, function, major equipment, and occupancy schedules.

Coal–Comm

Table 1 Commissioning team members and their primary responsibilities

Team member(s)	Primary responsibilities
Project manager	1. Coordinate the activities of building personnel and the commissioning team
	2. Schedule project activities
Continuous commissioning (CC) engineer(s)	1. Develop metering and field measurement plans
	2. Develop improved operational and control schedules
	3. Work with building staff to develop mutually acceptable implementation plans
	4. Make necessary programming changes to the building automation system
	5. Supervise technicians implementing mechanical systems changes
	6. Project potential performance changes and energy savings
	7. Conduct an engineering analysis of the system changes
	8. Write the project report
Designated facility staff	1. Participate in the initial facility survey
	2. Provide information about problems with facility operation
	3. Suggest commissioning measures for evaluation
	4. Approve all CC measures before implementation
	5. Actively participate in the implementation process
CC Technicians	1. Conduct field measurements
	2. Implement mechanical, electrical, and control system program modifications and changes, under the direction of the project engineer

- O&M records, if available.
- Description of any problems in the building, such as thermal comfort, indoor air quality, moisture, or mildew.

An experienced engineer should review this information and determine the potential of the CC process to improve comfort and reduce energy cost. If the CC potential is good, a CC assessment should be performed.

Step 2: *Perform CC Assessment and Develop Project Scope*. The CC assessment involves a site visit by an experienced commissioning engineer who examines EMCS screens, conducts spot measurements throughout the building systems, and identifies major CC measures suitable for the building. The CC assessment report lists and describes the preliminary CC measures identified, the estimated energy savings from implementation, and the cost of carrying out the CC process on the building(s) evaluated in the assessment. Once a commissioning contract is signed, the process moves to Phase 2.

Phase 2: CC Implementation and Verification

Step 1: *Develop CC Plan and Form the Project Team*. The CC project manager and project engineer develop a detailed work plan for the project that includes major tasks, their sequence, time requirements, and technical requirements. The work plan is then presented to the building owner or representative(s) at a meeting attended by any additional CC engineers and technicians on the project team. Owner contact personnel and in-house technicians who will work on the project are identified.

Step 2: *Develop Performance Baselines*. This step should document all known comfort problems in individual rooms resulting from too much heating, cooling, noise, humidity, or odors (especially from mold or mildew), or lack of outside air. Also, identify and document any HVAC system problems.

Baseline energy models of building performance are necessary to document the energy savings after commissioning. The baseline energy models can be developed using one or more of the following types of data:

- Short-term measured data obtained from data loggers or the EMCS system.
- Long-term hourly or 15-min whole building energy data, such as whole-building electricity, cooling, and heating consumption.
- Utility bills for electricity, gas, or chilled or hot water.

The baselines developed should be consistent with the International Performance Measurement and Verification Protocol,[7] with ASHRAE Guideline 14, or with both.[8]

Step 3: *Conduct System Measurements and Develop Proposed CC Measures*. The CC team uses EMCS trend data complemented by site measurements to identify current operational schedules and problems. The CC engineer conducts an engineering analysis to develop solutions for the existing problems; establishing improved operation and control schedules and set points for terminal boxes, air handling units (AHUs), exhaust systems, water and steam distribution systems, heat exchangers, chillers, boilers, and other components or systems as appropriate. Cost-effective energy retrofit measures can also be identified and documented during this step, if desired by the building owner.

Step 4: *Implement CC measures*. The CC project manager and/or project engineer presents the engineering solutions to existing problems and the improved operational and control schedules to the designated operating staff members and the building owner's representative to get "buy-in" and approval. Measures may be approved, modified, or rejected. A detailed implementation schedule is then developed by the CC engineer in consultation with the operating staff.

Continuous commissioning implementation normally starts by solving existing problems. Implementation of the improved operation and control schedules starts at the end of the comfort delivery system, such as at the terminal boxes, and ends with the central plant. The CC engineer closely supervises the implementation and refines the operational and control schedules as necessary. Following implementation, the new operation and control sequences are documented in a way that helps the building staff understand why they were implemented.

Step 5: *Document Comfort Improvements and Preliminary Energy Savings*. The comfort measurements taken in Step 2 (Phase 2) should be repeated at the same locations under comparable conditions and compared with the earlier measurements. The M&V procedures adopted in Step 2 should be used to determine the early post-CC energy performance and weather normalized to provide a preliminary evaluation of savings.

Step 6: *Keep the Commissioning Continuous*. The CC engineer should review the system operation after 6–12 months to identify any operating problems and make any adjustments needed. One year after CC implementation is complete, the CC engineer should write a project follow-up report that documents the first-year savings, recommendations or changes resulting from any consultation or site visits provided, and any recommendations to further improve building operations. Subsequently, the consumption should be tracked and compared with the first-year post-CC consumption during this period. Any significant and persistent increases in consumption should be investigated by the staff and/or CC engineer.

USES OF COMMISSIONING IN THE ENERGY MANAGEMENT PROCESS

Commissioning can be used as a part of the energy management program in several different ways:

- As a standalone measure. Commissioning is probably most often implemented in existing buildings because it is the most cost-effective step the owner can take to increase the energy efficiency of the building, generally offering a payback under three years, and often 1–2 years.
- As a follow-up to the retrofit process. Continuous commissioning has often been used to provide additional savings after a successful retrofit and has also been used numerous times to make an under-performing retrofit meet or exceed the original expectations.
- As an ECM in a retrofit program. The rapid payback that generally results from CC may be used to lower the payback of a package of measures to enable inclusion of a desired equipment replacement that has a longer payback in a retrofit package. This is illustrated by a case study in the next section. In this approach, the CC engineers conduct the CC audit in parallel with the retrofit audit conducted by the design engineering firm. Because the two approaches are different and look at different opportunities, it is very important to closely coordinate these two audits.
- To ensure that a new building meets or exceeds its energy performance goals. It may be used to significantly improve the efficiency of a new building by optimizing operation to meet its actual loads and uses instead of working to design assumptions.

CASE STUDY WITH CC AS AN ECM

Prairie View A&M University is a 1.7-million square foot campus, with most buildings served by a central thermal plant. Electricity is purchased from a local electric co-op.

University staff identified the need for major plant equipment replacements on campus. They wished to finance the upgrades through the Texas LoanSTAR program, which requires that the aggregate energy payback of all ECMs financed be ten years or less. Replacement of items such as chillers, cooling towers, and building automation systems typically have paybacks of considerably more than ten years. Hence, they can only be included in a loan if packaged with low payback measures that bring the aggregate payback below ten years.[9]

The university administration wanted to maximize the loan amount to get as much equipment replacement as possible. They also wanted to ensure that the retrofits worked properly after they are installed. To maximize their loan dollars, they chose to include CC as an ECM.

The LoanSTAR Program provides a brief walkthrough audit of the candidate buildings and plants. This audit is performed to determine whether there is sufficient retrofit potential to justify a more thorough investment grade audit.

The CC assessment is conducted in parallel with the retrofit audit conducted by the engineering design firm, when CC is to be included as an ECM. The two approaches look at different opportunities, but there can be some overlap, so it is very important to closely coordinate both audits. It is particularly important that the savings estimated by the audit team are not "double counted." The area of greatest overlap in this case was the building automation system. Considerable care was taken not to mix improved EMCS operation with operational improvements determined by the CC engineer, so both measures received proper credit.

The CC measures identified included the following:

- Hot and cold deck temperature resets.
- Extensive EMCS programming to avoid simultaneous heating and cooling.
- Air and water balancing.
- Duct static pressure resets.
- Sensor calibration and repair.
- Improved start, stop, warm-up, and shutdown schedules.

The CC engineers took the measurements required and collected adequate data on building operation during the CC assessment to perform a calibrated simulation on the major buildings. Available metered data and building EMCS data were also used. The CC energy savings were then written as an ECM and discussed with the design engineer. Any potential overlaps were removed. The combined ECMs were then listed and the total savings determined.

Table 2 summarizes the ECMs identified from the two audits:

The CC savings were calculated to be $204,563, as determined by conducting calibrated simulation of 16 campus buildings and by engineering calculations of savings from improved loop pumping. No CC savings were claimed for central plant optimization. Those savings were all applied to ECM #7, although it seems likely that additional CC savings will accrue from this measure. The simple payback from CC is slightly under three years, making it by far the most cost effective of the ECMs to be implemented. The CC savings represent nearly 30% of the total project savings.

Perhaps more importantly, CC accounted for two-thirds of the "surplus" savings dollars available to buy down the payback of the chillers and EMCS upgrade. Without CC as an ECM, the University would have had

Coal–Comm

Table 2 Summary of energy cost measures (ECMs)

ECM #	ECM	Annual savings				Cost to implement	Simple payback
		Electric kWh/yr	Electric demand kW/yr	Gas MCF/yr	Cost savings		
#1	Lighting	1,565,342	5221	(820)	$94,669	$561,301	6.0
#2	Replace chiller #3	596,891	1250	-0-	$33,707	$668,549	19.8
#3	Repair steam system	-0-	-0-	13,251	$58,616	$422,693	7.2
#4	Install motion sensors	81,616	-0-	(44.6)	$3567	$26,087	7.3
#5	Add 2 bldgs. to CW loop	557,676	7050	-0-	$60,903	$508,565	8.4
#6	Add chiller #4	599,891	1250	-0-	$33,707	$668,549	19.8
#7	Primary/ secondary pumping	1,070,207	-0-	-0-	$49,230	$441,880	9.0
#8	Replace DX systems	38,237	233	-0-	$2923	$37,929	13.0
#9	Replace DDC/ EMCS	2,969,962	670	2736	$151,488	$2,071,932	13.7
#10	Continuous commissioning	2,129,855	-0-	25,318	$204,563	$ 605,000	3.0
	Assessment reports					$102,775	
	Metering					$157,700	
	M&V					$197,500	
		9,606,677	15,674	40,440	$693,373	$6,470,460	9.3

to delete one chiller and the EMCS upgrades, or some combination of chillers and a portion of the building EMCS upgrades from the project to meet the ten-year payback criteria—one chiller and the EMCS upgrades, or some combination of chillers and limited building EMCS upgrades. With CC, however, the university was able to include all these hardware items, and still meet the ten-year payback.

SUMMARY

Commissioning of existing buildings is emerging as one of the most cost-effective ways for an energy manager to lower operating costs, and typically does so with no capital investment, or with a very minimal amount. It has been successfully implemented in several hundred buildings and provides typical paybacks of one to three years.

It is much more than the typical O&M program. It does not ensure that the systems function as originally designed, but focuses on improving overall system control and operations for the building as it is currently utilized and on meeting existing facility needs. During the CC process, a comprehensive engineering evaluation is conducted for both building functionality and system functions. The optimal operational parameters and schedules are developed based on actual building conditions. An integrated approach is used to implement these optimal schedules to ensure practical local and global system optimization and to ensure persistence of the improved operational schedules.

The approach presented in this chapter begins by conducting a thorough examination of all problem areas or operating problems in the building, diagnoses these problems, and develops solutions that solve these problems while almost always reducing operating costs at the same time. Equipment upgrades or retrofits may be implemented as well, but have not been a factor in the case studies presented, except where the commissioning was used to finance equipment upgrades. This is in sharp contrast to the more usual approach to improving the efficiency of HVAC systems and cutting operating costs, which primarily emphasizes system upgrades or retrofits to improve efficiency.

Commissioning of new buildings is also an important option for the energy manager, offering an opportunity to help ensure that new buildings have the energy efficiency and operational features that are most needed.

REFERENCES

1. ASHRAE. *ASHRAE Guideline 1–1996: The HVAC Commissioning Process*; American Society of Heating, Refrigerating and Air-Conditioning Engineers: Atlanta, GA, 1996.
2. U.S. Department of Energy. Building Commissioning: The Key to Quality Assurance, Washington, DC, 1999.
3. Continuous Commissioning® and CC® are registered trademarks of the Texas Engineering Experiment Station (TEES). Contact TEES for further information.
4. Claridge, D.E.; Turner, W.D.; Liu, M. et al. Is Commissioning Once Enough? Solutions for Energy Security & Facility Management Challenges: Proceedings of the 25th WEEC, Atlanta, GA, October 9–11, 2002, 29–36.
5. Mills, E.; Friedman, H.; Powell, T. et al. *The Cost-Effectiveness of Commercial-Buildings Commissioning: A Meta-Analysis of Energy and Non-Energy Impacts in Existing Buildings and New Construction in the United States*, Lawrence Berkeley National Laboratory Report No. 56637, 2004; 98. http://eetd.lbl.gov/EA/mills/emills.
6. Mills, E.; Bourassa, N.; Piette, M.A. et al. The Cost-Effectiveness of Commissioning New and Existing Commercial Buildings: Lessons from 224 Buildings. Proceedings of the 2006 National Conference on Building Commissioning, Portland Energy Conservation, Inc., New York, 2005. http://www.peci.org/ncbc/proceedings/2005/19_Piette_NCBC2005.pdf.
7. IPMVP Committee. International Performance Measurement & Verification Protocol: Concepts and Options for Determining Energy and Water Savings. U.S. Department of Energy, Vol. 1, DOE/GO-102001-1187, Washington, D.C., 2001.
8. ASHRAE Guideline 14-2002: Measurement of Energy and Demand Savings. American Society of Heating, Refrigerating and Air-Conditioning Engineers; Atlanta, GA, 2002.
9. Turner, W.D.; Claridge, D.E.; Deng, S.; Wei, G. The Use of Continuous Commissioning℠ as an Energy Conservation Measure (ECM) for Energy Efficiency Retrofits. Proceedings of 11th National Conference on Building Commisioning, Palm Springs, CA, CD, May 20–22, 2003.

BIBLIOGRAPHY

1. Two major sources of information on commissioning existing buildings are the *Continuous Commissioning℠ Guidebook: Maximizing Building Energy Efficiency and Comfort* (Liu, M., Claridge, D.E. and Turner, W.D., Federal Energy Management Program, U.S. Dept. of Energy, 144 pp., 2002) and *A Practical Guide for Commissioning Existing Buildings* (Haasl, T. and Sharp, T., Portland Energy Conservation, Inc. and Oak Ridge National Laboratory for U.S. DOE, ORNL/TM-1999/34, 69 pp. + App., 1999). Much of this entry has been abridged and adapted from the CC Guidebook.
2. The case studies in this entry have been largely abridged and adapted from Refs. 4 and 9.

Coal–Comm

Commissioning: New Buildings

Janey Kaster
Yamas Controls, Inc., South San Francisco, California, U.S.A.

Abstract

Commissioning is the methodology for bringing to light design errors, equipment malfunctions, and improper control strategies at the most cost-effective time to implement corrective action. The primary goal of commissioning is to achieve optimal building systems performance. There are two types of commissioning: acceptance-based and process-based. Process-based commissioning is a comprehensive process that begins in the predesign phase and continues through postacceptance, while acceptance-based commissioning, which is perceived to be the cheaper method, basically examines whether an installation is compliant with the design and accordingly achieves more limited results.

Commissioning originated in the early 1980s in response to a large increase in construction litigation. Commissioning was the result of owners seeking other means to gain assurance that they were receiving systems compliant with the design intent and with the performance characteristics and quality specified. Learn how commissioning has evolved and the major initiatives that are driving its growing acceptance.

The general rule for including a system in the commissioning process is: the more complicated the system is, the more compelling is the need to include it in the commissioning process. Other criteria for determining which systems should be included are discussed. Discover the many benefits of commissioning, such as improved quality assurance, dispute avoidance, and contract compliance.

Selection of the commissioning agent is key to the success of the commissioning plan. Learn what traits are necessary and what approaches to use for the selection process.

The commissioning process occurs over a variety of clearly delineated phases. The phases of the commissioning process as defined below are discussed in detail: predesign, design, construction/installation, acceptance, and postacceptance.

Extensive studies analyzing the cost/benefit of commissioning justify its application. One study defines the median commissioning cost for new construction as $1 per square foot or 0.6% of the total construction cost. The median simple payback for new construction projects utilizing commissioning is 4.8 years.

Understand how to achieve the benefits of commissioning, including optimization of building performance, reduction of facility life-cycle cost, and increased occupant satisfaction.

INTRODUCTION

This entry provides an overview of commissioning—the processes one employs to optimize the performance characteristics of a new facility being constructed. Commissioning is important to achieve customer satisfaction, optimal performance of building systems, cost containment, and energy efficiency, and it should be understood by contractors and owners.

After providing an overview of commissioning and its history and prevalence, this entry discusses what systems should be part of the commissioning process, the benefits of commissioning, how commissioning is conducted, and the individuals and teams critical for successful commissioning. Then the entry provides a detailed discussion of each of the different phases of a successful commissioning process, followed by a discussion of the common mistakes to avoid and how one can measure the success of a commissioning effort, together with a cost–benefit analysis tool.

The purpose of this entry will be realized if its readers decide that successful commissioning is one of the most important aspects of construction projects and that commissioning should be managed carefully and deliberately throughout any project, from predesign to postacceptance. As an introduction to those unfamiliar with the process and as a refresher for those who are, the following section provides an overview of commissioning, how it developed, and its current prevalence today.

OVERVIEW OF COMMISSIONING

Commissioning Defined

Commissioning is the methodology for bringing to light design errors, equipment malfunctions, and improper control strategies at the most cost-effective time to implement corrective action. Commissioning facilitates a thorough understanding of a facility's intended use and

Keywords: Commissioning; New building system optimization; Construction process; Functional testing.

ensures that the design meets the intent through coordination, communication, and cooperation of the design and installation team. Commissioning ensures that individual components function as a cohesive system. For these reasons, commissioning is best when it begins in the predesign phase of a construction project and can in one sense be viewed as the most important form of quality assurance for construction projects.

Unfortunately, there are many misconceptions associated with commissioning, and perhaps for this reason, commissioning has been executed with varying degrees of success, depending on the level of understanding of what constitutes a "commissioned" project. American Society of Heating, Refrigerating and Air-Conditioning Engineers (AHSRAE) guidelines define commissioning as: the process of ensuring that systems are designed, installed, functionally tested, and capable of being operated and maintained to perform conformity with the design intent… [which] begins with planning and includes design, construction, startup, acceptance, and training, and is applied throughout the life of the building.[4] However, for many contractors and owners, this definition is simplified into the process of system startup and checkout or completing punch-list items.

Of course, a system startup and checkout process carried out by a qualified contractor is one important aspect of commissioning. Likewise, construction inspection and the generation and completion of punch-list items by a construction manager are other important aspects of commissioning. However, it takes much more than these standard installation activities to have a truly "commissioned" system. Commissioning is a comprehensive and methodical approach to the design and implementation of a cohesive system that culminates in the successful turnover of the facility to maintenance staff trained in the optimal operation of those systems.

Without commissioning, a contractor starts up the equipment but doesn't look beyond the startup to system operation. Assessing system operation requires the contractor to think about how the equipment will be used under different conditions. As one easily comprehended example, commissioning requires the contractor to think about how the equipment will operate as the seasons change. Analysis of the equipment and building systems under different load conditions due to seasonal conditions at the time of system startup will almost certainly result in some adjustments to the installed equipment for all but the most benign climates. However, addressing this common requirement of varying load due to seasonal changes most likely will not occur without commissioning. Instead, the maintenance staff is simply handed a building with minimal training and left to figure out how to achieve optimal operation on their own. In this seasonal example, one can just imagine how pleased the maintenance staff would be with the contractor when a varying load leads to equipment or system failure—often under very hot or very cold conditions!

Thus, the primary goal of commissioning is to achieve optimal building systems performance. For heating, ventilation, and air-conditioning (HVAC) systems, optimal performance can be measured by thermal comfort, indoor air quality, and energy savings. Energy savings, however, can result simply from successful commissioning targeted at achieving thermal comfort and excellent indoor air quality. Proper commissioning will prevent HVAC system malfunction—such as simultaneous heating and cooling, and overheating or overcooling—and successful malfunction prevention translates directly into energy savings. Accordingly, energy savings rise with increasing comprehensiveness of the commissioning plan. Commissioning enhances energy performance (savings) by ensuring and maximizing the performance of specific energy efficiency measures and correcting problems causing excessive energy use.[3] Commissioning, then, is the most cost-effective means of improving energy efficiency in commercial buildings. In the next section, the two main types of commissioning in use today—acceptance-based and processed-based—are compared and contrasted.

Acceptance-Based vs Process-Based Commissioning

Given the varied nature of construction projects, contractors, owners, buildings, and the needs of the diverse participants in any building projection, commissioning can of course take a variety of forms. Generally, however, there are two types of commissioning: acceptance-based and process-based. Process-based commissioning is a comprehensive process that begins in the predesign phase and continues through postacceptance, while acceptance-based commissioning, which is perceived to be the cheaper method, basically examines whether an installation is compliant with the design and accordingly achieves more limited results.

Acceptance-based commissioning is the most prevalent type due to budget constraints and the lack of hard cost/benefit data to justify the more extensive process-based commissioning. Acceptance-based commissioning does not involve the contractor in the design process but simply constitutes a process to ensure that the installation matches the design. In acceptance-based commissioning, confrontational relationships are more likely to develop between the commissioning agent and the contractor because the commissioning agent and the contractor, having been excluded from the design phase, have not "bought in" to the design and thus may be more likely to disagree in their interpretation of the design intent.

Because the acceptance-based commissioning process simply validates that the installation matches the design, installation issues are identified later in the cycle.

Construction inspection and regular commissioning meetings do not occur until late in the construction/installation phase with acceptance-based commissioning. As a result, there is no early opportunity to spot errors and omissions in the design, when remedial measures are less costly to undertake and less likely to cause embarrassment to the designer and additional costs to the contractor. As most contractors will readily agree, addressing issues spotted in the design or submittal stages of construction is typically much less costly than addressing them after installation, when correction often means tearing out work completed and typically delays the completion date.

Acceptance-based commissioning is cheaper, however, at least on its face, being approximately 80% of the cost of process-based commissioning.[2] If only the initial cost of commissioning services is considered, many owners will conclude that this is the most cost-effective commissioning approach. However, this 20% cost differential does not take into account the cost of correcting defects after the fact that process-based commissioning could have identified and corrected at earlier stages of the project. One need encounter only a single, expensive-to-correct project to become a devotee of process-based commissioning.

Process-based commissioning involves the commissioning agent in the predesign through the construction, functional testing, and owner training. The main purpose is quality assurance—assurance that the design intent is properly defined and followed through in all phases of the facility life cycle. It includes ensuring that the budget matches the standards that have been set forth for the project so that last-minute "value engineering" does not undermine the design intent, that the products furnished and installed meet the performance requirements and expectation compliant with the design intent, and that the training and documentation provided to the facility staff equip them to maintain facility systems true to the design intent.

As the reader will no doubt already appreciate, the author believes that process-based commissioning is far more valuable to contractors and owners than acceptance-based commissioning. Accordingly, the remainder of this entry will focus on process-based commissioning, after a brief review of the history of commissioning from inception to date, which demonstrates that our current, actively evolving construction market demands contractors and contracting professionals intimately familiar with and expert in conducting process-based commissioning.

History of Commissioning

Commissioning originated in the early 1980s in response to a large increase in construction litigation. Owners were dissatisfied with the results of their construction projects and had recourse only to the courts and litigation to resolve disputes that could not be resolved by meeting directly with their contractors. While litigation attorneys no doubt found this satisfactory approach to resolving construction project issues, owners did not, and they actively began looking for other means to gain assurance that they were receiving systems compliant with the design intent and with the performance characteristics and quality specified. Commissioning was the result.

While commissioning enjoyed early favor and wide acceptance, the recession of the mid-1980s placed increasing market pressure on costs, and by the mid-to late 1980s it forced building professionals to reduce fees and streamline services. As a result, acceptance-based commissioning became the norm, and process-based commissioning became very rare. This situation exists in most markets today; however, the increasing cost of energy, the growing awareness of the global threat of climate change and the need to reduce CO_2 emissions as a result, and the legal and regulatory changes resulting from both are creating a completely new market in which process-based commissioning will become ever more important, as discussed in the following section.

Prevalence of Commissioning Today

There are varying degrees of market acceptance of commissioning from state to state. Commissioning is in wide use in California and Texas, for example, but it is much less widely used in many other states. The factors that impact the level of market acceptance depend upon:

- The availability of commissioning service providers
- State codes and regulations
- Tax credits
- Strength of the state's economy[1]

State and federal policies with regard to commissioning are changing rapidly to increase the demand for commissioning. Also, technical assistance and funding are increasingly available for projects that can serve as demonstration projects for energy advocacy groups. The owner should investigate how each of these factors could benefit the decision to adopt commissioning in future construction projects.

Some of the major initiatives driving the growing market acceptance of commissioning are:

- Federal government's U.S. Energy Policy Act of 1992 and Executive Order 12902, mandating that federal agencies develop commissioning plans
- Portland Energy Conservation, Inc.; National Strategy for Building Commissioning; and their annual conferences
- ASHRAE HVAC Commissioning Guidelines (1989)
- Utilities establishing commissioning incentive programs
- Energy Star building program
- Leadership in Energy Environmental Design (LEED) certification for new construction

- Building codes
- State energy commission research programs

Currently, the LEED is having the largest impact in broadening the acceptance of commissioning. The Green Building Council is the sponsor of LEED and is focused on sustainable design—design and construction practices that significantly reduce or eliminate the cradle-to-grave negative impacts of buildings on the environment and building occupants. Leadership in energy efficient design encourages sustainable site planning, conservation of water and water efficiency, energy efficiency and renewable energy, conservation of materials and resources, and indoor environmental quality.

With this background on commissioning, the various components of the commissioning process can be explored, beginning with an evaluation of what building systems should be subject to the commissioning process.[5]

COMMISSIONING PROCESS

Systems to Include in the Commissioning Process

The general rule for including a system in the commissioning process is: the more complicated the system is the more compelling is the need to include it in the commissioning process. Systems that are required to integrate or interact with other systems should be included. Systems that require specialized trades working independently to create a cohesive system should be included, as well as systems that are critical to the operation of the building. Without a commissioning plan on the design and construction of these systems, installation deficiencies are likely to create improper interaction and operation of system components.

For example, in designing a lab, the doors should be included in the commissioning process because determining the amount of leakage through the doorways could prove critical to the ability to maintain critical room pressures to ensure proper containment of hazardous material. Another common example is an energy retrofit project. Such projects generally incorporate commissioning as part of the measurement and verification plan to ensure that energy savings result from the retrofit process.

For any project, the owner must be able to answer the question of why commissioning is important.

Why Commissioning?

A strong commissioning plan provides quality assurance, prevents disputes, and ensures contract compliance to deliver the intended system performance. Commissioning is especially important for HVAC systems that are present in virtually all buildings because commissioned HVAC systems are more energy efficient.

The infusion of electronics into almost every aspect of modern building systems creates increasingly complex systems requiring many specialty contractors. Commissioning ensures that these complex subsystems will interact as a cohesive system.

Commissioning identifies design or construction issues and, if done correctly, identifies them at the earliest stage in which they can be addressed most cost effectively. The number of deficiencies in new construction exceeds existing building retrofit by a factor of 3.[3] Common issues that can be identified by commissioning that might otherwise be overlooked in the construction and acceptance phase are: air distribution problems (these occur frequently in new buildings due to design capacities, change of space utilization, or improper installation), energy problems, and moisture problems.

Despite the advantages of commissioning, the current marketplace still exhibits many barriers to adopting commissioning in its most comprehensive and valuable forms.

Barriers to Commissioning

The general misperception that creates a barrier to the adoption of commissioning is that it adds extra, unjustified costs to a construction project. Until recently, this has been a difficult perception to combat because there are no energy-use baselines for assessing the efficiency of a new building. As the cost of energy continues to rise, however, it becomes increasingly less difficult to convince owners that commissioning is cost effective. Likewise, many owners and contractors do not appreciate that commissioning can reduce the number and cost of change orders through early problem identification. However, once the contractor and owner have a basis on which to compare the benefit of resolving a construction issue earlier as opposed to later, in the construction process, commissioning becomes easier to sell as a win–win proposal.

Finding qualified commissioning service providers can also be a barrier, especially in states where commissioning is not prevalent today. The references cited in this entry provide a variety of sources for identifying associations promulgating commissioning that can provide referrals to qualified commissioning agents.

For any owner adopting commissioning, it is critical to ensure acceptance of commissioning by all of the design construction team members. Enthusiastic acceptance of commissioning by the design team will have a very positive influence on the cost and success of your project. An objective of this entry is to provide a source of information to help gain such acceptance by design construction team members and the participants in the construction market.

Selecting the Commissioning Agent

Contracting an independent agent to act on behalf of the owner to perform the commissioning process is the best way to ensure successful commissioning. Most equipment vendors are not qualified and are likely to be biased against discovering design and installation problems—a critical function of the commissioning agent—with potentially costly remedies. Likewise, systems integrators have the background in control systems and data exchange required for commissioning but may not be strong in mechanical design, which is an important skill for the commissioning agent. Fortunately, most large mechanical consulting firms offer comprehensive commissioning services, although the desire to be competitive in the selection processes sometimes forces these firms to streamline their scope on commissioning.

Owners need to look closely at the commissioning scope being offered. An owner may want to solicit commissioning services independently from the selection of the architect/mechanical/electrical/plumbing design team or, minimally, to request specific details on the design team's approach to commissioning. If an owner chooses the same mechanical, electrical, and plumbing (MEP) firm for design and commissioning, the owner should ensure that there is physical separation between the designer and commissioner to ensure that objectivity is maintained in the design review stages. An owner should consider taking on the role of the commissioning agent directly, especially if qualified personnel exist in-house. This approach can be very cost effective. The largest obstacles to success with an in-house commissioning agent are the required qualifications and the need to dedicate a valuable resource to the commissioning effort. Many times, other priorities may interfere with the execution of the commissioning process by an in-house owner's agent.

There are three basic approaches to selecting the commissioning agent:

Negotiated—best approach for ensuring a true partnership
Selective bid list—preapproved list of bidders
Competitive—open bid list

Regardless of the approach, the owner should clearly define the responsibilities of the commissioning agent at the start of the selection process. Fixed-cost budgets should be provided by the commissioning agent to the owner for the predesign and design phases of the project, with not-to-exceed budgets submitted for the construction and acceptance phases. Firm service fees should be agreed upon as the design is finalized.

Skills of a Qualified Commissioning Agent

A commissioning agent needs to be a good communicator, both in writing and verbally. Writing skills are important because documentation is critical to the success of the commissioning plan. Likewise, oral communication skills are important because communicating issues uncovered in a factual and nonaccusatory manner is most likely to resolve those issues efficiently and effectively. The commissioning agent should have practical field experience in MEP controls design and startup to be able to identify potential issues early. The commissioning agent likewise needs a thorough understanding of how building structural design impacts building systems. The commissioning agent must be an effective facilitator and must be able to decrease the stress in stressful situations. In sum, the commissioning agent is the cornerstone of the commissioning team and the primary determinant of success in the commissioning process.

At least ten organizations offer certifications for commissioning agents. However, there currently is no industry standard for certifying a commissioning agent. Regardless of certification, the owner should carefully evaluate the individuals to be performing the work from the commissioning firm selected. Individual experience and reputation should be investigated. References for the lead commissioning agent are far more valuable than references for the executive members of a commissioning firm in evaluating potential commissioning agents. The commissioning agent selected will, however, only be one member of a commissioning team, and the membership of the commissioning team is critical to successful commissioning.

Commissioning Team

The commissioning team is composed of representatives from all members of the project delivery team: the commissioning agent, representatives of the owner's maintenance team, the architect, the MEP designer, the construction manager, and systems contractors. Each team member is responsible for a particular area of expertise, and one important function of the commissioning agent is to act as a facilitator of intrateam communication.

The maintenance team representatives bring to the commissioning team the knowledge of current operations, and they should be involved in the commissioning process at the earliest stage, defining the design intent in the predesign phase, as described below. Early involvement of maintenance team representatives ensures a smooth transition from construction to a fully operational facility, and aids in the acceptance and full use of the technologies and strategies that have been developed during the commissioning process. Involvement of the maintenance team representatives also shortens the building turnover transition period.

The other members of the commissioning team have defined and important functions. The architect leads the development of the design intent document (DID). The MEP designer's responsibilities are to develop the mechanical systems that support the design intent of the facility and comply with the owner's current operating standards. The MEP schematic design is the basis for the systems installed and is discussed further below. The construction manager ensures that the project installation meets the criteria defined in the specifications, the budget requirements, and the predefined schedule. The systems contractors' responsibilities are to furnish and install a fully functional system that meets the design specifications. There are generally several contractors whose work must be coordinated to ensure that the end product is a cohesive system.

Once the commissioning team is in place, commissioning can take place, and it occurs in defined and delineated phases—the subject of the following section.

COMMISSIONING PHASES

The commissioning process occurs over a variety of clearly delineated phases. The commission plan is the set of documents and events that defines the commissioning process over all phases. The commissioning plan needs to reflect a systematic, proactive approach that facilitates communication and cooperation of the entire design and construction team.

The phases of the commissioning process are:

Predesign
Design
Construction/installation
Acceptance
Postacceptance

These phases and the commissioning activities associated with them are described in the following sections.

Predesign Phase

The predesign phase is the phase in which the design intent is established in the form of the DID. In this phase of a construction project, the role of commissioning in the project is established if process-based commissioning is followed. Initiation of the commissioning process in the predesign phase increases acceptance of the commissioning process by all design team members. Predesign discussions about commissioning allow all team members involved in the project to assess and accept the importance of commissioning to a successful project. In addition, these discussions give team members more time to assimilate the impact of commissioning on their individual roles and responsibilities in the project. A successful project is more likely to result when the predesign phase is built around the concept of commissioning instead of commissioning's being imposed on a project after it has been designed.

Once an owner has decided to adopt commissioning as an integral part of the design and construction of a project, the owner should be urged to follow the LEED certification process, as discussed above. The commissioning agent can assist in the documentation preparation required for the LEED certification, which occurs in the postacceptance phase.

The predesign phase is the ideal time for an owner to select and retain the commissioning agent. The design team member should, if possible, be involved in the selection of the commissioning agent because that member's involvement will typically ensure a more cohesive commissioning team. Once the commissioning agent is selected and retained, the commissioning-approach outline is developed. The commissioning-approach outline defines the scope and depth of the commissioning process to be employed for the project. Critical commissioning questions are addressed in this outline. The outline will include, for most projects, answers to the following questions:

What equipment is to be included?
What procedures are to be followed?
What is the budget for the process?

As the above questions suggest, the commissioning budget is developed from the choices made in this phase. Also, if the owner has a commissioning policy, it needs to be applied to the specifics of the particular project in this phase.

The key event in the predesign phase is the creation of the DID, which defines the technical criteria for meeting the requirements of the intended use of the facilities. The DID document is often created based in part upon the information received from interviews with the intended building occupants and maintenance staff. Critical information—such as the hours of operation, occupancy levels, special environmental considerations (such as pressure and humidity), applicable codes, and budgetary considerations and limitations—is identified in this document. The owner's preference, if any, for certain equipment or contactors should also be identified at this time. Together, the answers to the critical questions above and the information in the DID are used to develop the commissioning approach outline. A thorough review of the DID by the commissioning agent ensures that the commissioning-approach outline will be aligned with the design intent.

With the commissioning agent selected, the DID document created, and the commissioning approach outline in place, the design phase is ready to commence.

Design Phase

The design phase is the phase in which the schematics and specifications for all components of a project are prepared.

One key schematic and set of specifications relevant to the commissioning plan is the MEP schematic design, which specifies installation requirements for the MEP systems. As noted, the DID is the basis for creating the commissioning approach outline in the predesign phase. The DID also serves as the basis for creating the MEP schematic design in the design phase. The DID provides the MEP designer with the key concepts from which the MEP schematic design is developed.

The completed MEP schematic design is reviewed by the commissioning agent for completeness and conformance to the DID. At this stage, the commissioning agent and the other design team members should consider what current technologies, particularly those for energy efficiency, could be profitably included in the design. Many of the design enhancements currently incorporated into existing buildings during energy retrofitting for operational optimization are often not considered in new building construction. This can result in significant lost opportunity, so these design enhancements should be reviewed for incorporation into the base design during this phase of the commissioning process. This point illustrates the important principle that technologies important to retrocommissioning should be applied to new building construction—a point that is surprisingly often overlooked in the industry today.

For example, the following design improvements and technologies should always be considered for applicability to a particular project:

- Variable-speed fan and pumps installed
- Chilled water cooling (instead of DX cooling)
- Utility meters for gas, electric, hot water, chilled water, and steam at both the building and system level
- CO_2 implementation for minimum indoor air requirements

This list of design improvements is not exhaustive; the skilled commissioning agent will create and expand personalized lists as experience warrants and as the demands of particular projects suggest.

In addition to assisting in the evaluation of potential design improvements, the commissioning agent further inspects the MEP schematic design for:

- Proper sizing of equipment capacities
- Clearly defined and optimized operating sequences
- Equipment accessibility for ease of servicing

Once the commissioning agent's review is complete, the feedback is discussed with the design team to determine whether its incorporation into the MEP schematic design is warranted. The agreed-upon changes or enhancements are incorporated, thus completing the MEP schematic design.

The completed MEP schematic design serves as the basis on which the commissioning agent will transform the commissioning-approach outline into the commissioning specification.

The commissioning specification is the mechanism for binding contractually the contractors to the commissioning process. Expectations are clearly defined, including:

- Responsibilities of each contractor
- Site meeting requirements
- List of the equipment, systems, and interfaces
- Preliminary verification checklists
- Preliminary functional-performance testing checklists
- Training requirements and who is to participate
- Documentation requirements
- Postconstruction documentation requirements
- Commissioning schedule
- Definition for system acceptance
- Impact of failed results

Completion of the commissioning specification is required to select the systems contractor in a competitive solicitation. Alternatively, however, owners with strong, preexisting relationships with systems contractors may enter into a negotiated bid with those contractors, who can then be instrumental in finalizing the commissioning specification.

Owners frequently select systems contractors early in the design cycle to ensure that the contractors are involved in the design process. As noted above, if there are strong, preexisting relationships with systems contractors, early selection without a competitive selection process (described in the following paragraph) can be very beneficial. However, if there is no competitive selection process, steps should be taken to ensure that the owner gets the best value. For example, unit pricing should be negotiated in advance to ensure that the owner is getting fair and reasonable pricing. The commissioning agent and the MEP designer can be good sources for validating the unit pricing. The final contract price should be justified with the unit pricing information.

If the system selection process is competitive, technical proposals should be requested with the submission of the bid price. The systems contractors need to demonstrate a complete understanding of the project requirements to ensure that major components have not been overlooked. Information such as the project schedule and manpower loading for the project provide a good basis from which to measure the contractor's level of understanding. If the solicitation does not have a preselected list of contractors, the technical proposal should include the contractor's financial information, capabilities, and reference lists. As in the negotiated process described above, unit pricing should be requested to ensure the proper pricing of project additions and deletions. The review of the technical proposals should be included in the commissioning agent's scope of work.

A mandatory prebid conference should be held to walk the potential contractors through the requirements and to reinforce expectations. This conference should be held regardless of the approach—negotiated or competitive—used for contractor selection. The contractor who is to bear the financial burden for failed verification tests and subsequent functional-performance tests should be reminded of these responsibilities to reinforce their importance in the prebid meeting. The prebid conference sets the tone of the project and emphasizes the importance of the commissioning process to a successful project.

Once the MEP schematic design and commissioning specification are complete, and the systems contractors have been selected, the construction/installation phase begins.

Construction/Installation Phase

Coordination, communication, and cooperation are the keys to success in the construction and installation phase. The commissioning agent is the catalyst for ensuring that these critical activities occur throughout the construction and installation phase.

Frequently, value engineering options are proposed by the contractors prior to commencing the installation. The commissioning agent should be actively involved in the assessment of any options proposed. Many times, what appears to be a good idea in construction can have a disastrous effect on a facility's long-term operation. For example, automatic controls are often value engineered out of the design, yet the cost of their inclusion is incurred many times over in the labor required to perform their function manually over the life of the building. The commissioning agent can ensure that the design intent is preserved, the life-cycle costs are considered, and the impact on all systems of any value engineering modification proposed is thoroughly evaluated.

Once the design aspects are complete and value engineering ideas have been incorporated or rejected, the submittals, including verification checklists, need to be finalized. The submittals documentation is prepared by the systems contractors and reviewed by the commissioning agent. There are two types of submittals: technical submittals and commissioning submittals. Both types of submittals are discussed below.

Technical submittals are provided to document the systems contractors' interpretation of the design documents. The commissioning agent reviews the technical submittals for compliance and completeness. It is in this submittal review process that potential issues are identified prior to installation, reducing the need for rework and minimizing schedule delays. The technical submittals should include:

- Detailed schematics
- Equipment data sheets
- Sequence of operation
- Bill of material

A key technical submittal is the testing, adjusting, and balancing submittal (TAB). The TAB should include:

- TAB procedures
- Instrumentation
- Format for results
- Data sheets with equipment design parameters
- Operational readiness requirements
- Schedule

In addition to the TAB, other technical submittals, such as building automation control submittals, will be obtained from the systems contractors and reviewed by the commissioning agent.

The commissioning submittal generally follows the technical submittal in time and includes:

- Verification checklists
- Startup requirements
- Test and balance plan
- Training plan

The commissioning information in the commissioning submittal is customized for each element of the system.

These submittals, together with the commissioning specification, are incorporated into the commissioning plan, which becomes a living document codifying the results of the construction commissioning activities. This plan should be inspected in regular site meetings. Emphasis on the documentation aspect of the commissioning process early in the construction phase increases the contractors' awareness of the importance of commissioning to a successful project.

In addition to the submittals, the contractors are responsible for updating the design documents with submitted and approved equipment data and field changes on an ongoing basis. This update design document should be utilized during the testing and acceptance phase.

The commissioning agent also performs periodic site visits during the installation to observe the quality of workmanship and compliance with the specifications. Observed deficiencies should be discussed with the contractor and documented to ensure future compliance. Further inspections should be conducted to ensure that appropriate corrective action has been taken.

The best way to ensure that the items discussed above are addressed in a timely manner is to hold regularly scheduled commissioning meetings that require the participation of all systems contractors. This is the mechanism for ensuring that communication occurs. Meeting minutes prepared by the commissioning agent document the discussions and decisions reached. Commissioning meetings should be coordinated with the regular project meetings because many participants in a construction project need to attend both meetings.

Typical elements of a commissioning meeting include:

- Discussing field installation issues to facilitate rapid response to field questions
- Updating design documents with field changes
- Reviewing the commissioning agent's field observations
- Reviewing progress against schedule
- Coordinating multicontractor activities

Once familiar with the meeting process, an agenda will be helpful but not necessary. Meeting minutes should be kept and distributed to all participants.

With approved technical and commissioning plan submittals, as installation progresses, the contractor is ready to begin the system verification testing. The systems contractor generally executes the system verification independently of the commissioning agent. Contractor system verification includes:

- Point-to-point wiring checked out
- Sensor accuracy validated
- Control loops exercised

Each of the activities should be documented for each control or system element, and signed and dated by the verification technician.

The documentation expected from these activities should be clearly defined in the commissioning specification to ensure its availability to the commissioning agent for inspection of the verification process. The commissioning agent's role in the system verification testing is to ensure that the tests are completed and that the results reflect that the system is ready for the functional-performance tests. Because the commissioning agent is typically not present during the verification testing, the documentation controls how successfully the commissioning agent performs this aspect of commissioning.

In addition to system verification testing, equipment startup is an important activity during this phase. Equipment startup occurs at different time frames relative to the system verification testing, depending on the equipment and system involved. There may be instances when the system verification needs to occur prior to equipment startup to prevent a catastrophic event that could lead to equipment failure. The commissioning agent reviews the startup procedures prior to the startup to ensure that equipment startup is coordinated properly with the system verification. Unlike in verification testing, the commissioning agent should be present during HVAC equipment startup to document the results. These results are memorialized in the final commissioning report, so their documentation ultimately is the responsibility of the commissioning agent.

Once system verification testing and equipment startup have been completed, the acceptance phase begins.

Acceptance Phase

The acceptance phase of the project is the phase in which the owner accepts the project as complete and delivered in accordance with the specifications, and concludes with acceptance of the project in its entirety. An effective commissioning process during the installation phase should reduce the time and labor associated with the functional-performance tests of the acceptance phase.

Statistical sampling is often used instead of 100% functional-performance testing to make the process more efficient. A 20% random sample with a failure rate less than 1% indicates that the entire system was properly installed. If the failure rate exceeds 1%, a complete testing of every system may need to be completed to correct inadequacies in the initial checkout and verification testing. This random-sampling statistical approach holds the contractor accountable for the initial checkout and test, with the performance testing serving only to confirm the quality and thoroughness of the installation. This approach saves time and money for all involved. It is critical, however, that the ramifications of not meeting the desired results of the random tests are clearly defined in the commissioning specifications.

The commissioning agent witnesses and documents the results of the functional-performance tests, using specific forms and procedures developed for the system being tested. These forms are created with the input of the contractor in the installation phase. Involvement of the maintenance staff in the functional-performance testing is important. The maintenance team is often not included in the design process, so they may not fully understand the design intent. The functional-performance testing can provide the maintenance team an opportunity to learn and appreciate the design intent. If the design intent is to be preserved, the maintenance team must fully understand the design intent. This involvement of the maintenance team increases their knowledge of the system going into the training and will increase the effectiveness of the training.

Training of the maintenance team is critical to a successful operational handover once a facility is ready for occupancy. This training should include:

- Operations and maintenance (O&M) manual overview
- Hardware component review
- Software component review
- Operations review
- Interdependencies discussion
- Limitations discussion
- Maintenance review
- Troubleshooting procedures review
- Emergency shutdown procedures review

The support level purchased from the systems contractor determines the areas of most importance in the training and therefore should be determined prior to the training process. Training should be videotaped for later use by new maintenance team members and in refresher courses, and for general reference by the existing maintenance team. Using the O&M manuals as a training manual increases the maintenance team's awareness of the information contained in them, making the O&M manuals more likely to be referenced when appropriate in the future.

The O&M manuals should be prepared by the contractor in an organized and easy-to-use manner. The commissioning agent is sometimes engaged to organize them all into an easily referenced set of documents. The manuals should be provided in both hard-copy and electronic formats, and should include:

- System diagrams
- Input/output lists
- Sequence of operations
- Alarm points list
- Trend points list
- Testing documentation
- Emergency procedures

These services—including functional-performance testing, training, and preparing O&M manuals—should be included in the commissioning plan to ensure the project's successful acceptance. The long-term success of the project, however, is determined by the activities that occur in the postacceptance phase.

Postacceptance Phase

The postacceptance phase is the phase in which the owner takes beneficial occupancy and forms an opinion about future work with the design team, contractors, and the commissioning agent who completed the project. This is also the phase in which LEED certification, if adopted, is completed. Activities that usually occur in the acceptance phase should instead occur in the postacceptance phase. This is due to constraints that are not controllable by the contractor or owner. For example, seasonal changes may make functional-performance testing of some HVAC systems impractical during the acceptance phase for certain load conditions. This generally means that in locations that experience significant seasonal climate change, some of the functional-performance testing is deferred until suitable weather conditions exist. The commissioning agent determines which functional-performance tests need to be deferred and hence carried out in the postacceptance phase.

During the postacceptance phase, the commissioning agent prepares a final commissioning report that is provided to the owner and design team. The executive summary of this report provides an overall assessment of the design intent conformance. The report details whether the commissioned equipment and systems meet the commissioning requirements. Problems encountered and corrective actions taken are documented in this report. The report also includes the signed and dated startup and functional-performance testing checklists.

The final commissioning report can be used profitably as the basis of a "lessons learned" meeting involving the design team so that the commissioning process can be continuously improved and adaptations can be made to the owner's commissioning policy for future projects. The owner should use the experience of the first commissioned project to develop the protocols and standards for future projects. The documentation of this experience is the owner's commissioning policy. Providing this policy and the information it contains to the design and construction team for the next project can help the owner reduce budget overruns by eliminating any need to reinvent protocols and standards and by setting the right expectations earlier in the process.

Commissioning therefore should not be viewed as a one-time event but should instead be viewed as an operational philosophy. A recommissioning or continuous commissioning plan should be adopted for any building to sustain the benefits delivered from a commissioning plan. The commissioning agent can add great value to the creation of the recommissioning plan and can do so most effectively in the postacceptance phase of the project.

Fig. 1 depicts the information development that occurs in the evolution of a commissioning plan and summarizes the information presented in the preceding sections by outlining the various phases of the commissioning process.

With this background, the reader is better positioned for success in future commissioning projects and better prepared to learn the key success factors in commissioning and how to avoid common mistakes in the commissioning process.

COMMISSIONING SUCCESS FACTORS

Ultimately, the owner will be the sole judge of whether a commissioning process has been successful. Thus, second only to the need for a competent, professional commissioning agent, keeping the owner or the owner's senior representative actively involved in and informed at all steps of the commissioning process is a key success factor. The commissioning agent should report directly to the owner or the owner's most senior representative on the project, not only to ensure that this involvement and information transfer occur, but also to ensure the objective implementation of the commissioning plan—a third key success factor.

Another key success factor is an owner appreciation—which can be enhanced by the commissioning agent—that commissioning must be an ongoing process to get full

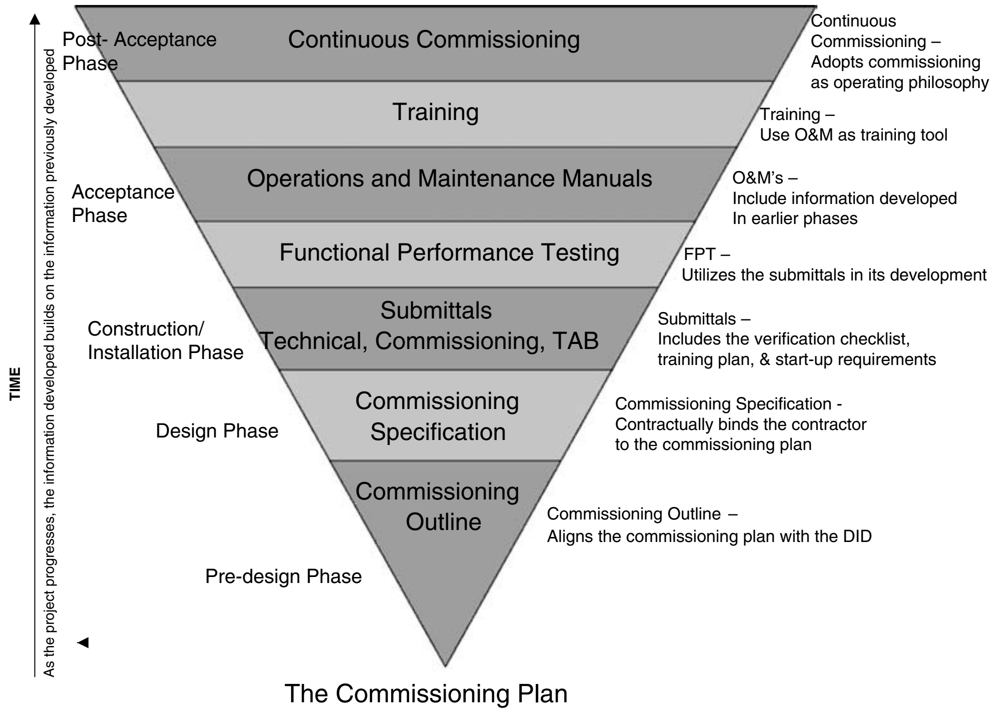


Fig. 1 The commissioning plan.

benefit. For example, major systems should undergo periodic modified functional testing to ensure that the original design intent is being maintained or to make system modification if the design intent has changed. If an owner appreciates that commissioning is a continuous process that lasts for the entire life of the facility, the commissioning process will be a success.

Most owners will agree that the commissioning process is successful if success can be measured in a cost/benefit analysis. Cost/benefit or return on equity is the most widely used approach to judge the success of any project. Unfortunately, the misapplication of cost/benefit analyses has been the single largest barrier to the widespread adoption of commissioning. For example, because new construction does not have an energy baseline from which to judge energy savings, improper application of a cost/benefit analysis can lead to failure to include energy savings technologies—technologies the commissioning agent can identify—in the construction process. Similarly, unless one can appreciate how commissioning can prevent schedule delays and rework by spotting issues and resolving them early in the construction process, one cannot properly offset the costs of commissioning with the benefits.

Fortunately, there are now extensive studies analyzing the cost/benefit of commissioning that justify its application. A study performed jointly by Lawrence Berkeley National Laboratory; Portland Energy Conservation, Inc.; and the Energy Systems Laboratory at Texas A&M University provides compelling analytical data on the cost/benefit of commissioning. The study defines the median commissioning cost for new construction as $1 per square foot or 0.6% of the total construction cost. The median simple payback for new construction projects utilizing commissioning is 4.8 years. This simple payback calculation does not take into account the quantified nonenergy impacts, such as the reduction in the cost and frequency of change orders or premature equipment failure due to improper installation practices. The study quantifies the median nonenergy benefits for new construction at $1.24 per square foot per year.[3]

While the primary cost component of assessing the cost/benefit of commissioning lies in whether there was a successful negotiation of the cost of services with the commissioning service provider, the more important aspect of the analysis relates to the outcomes of the process. For example, after a commissioning process is complete, what are the answers to these questions?

Are the systems functioning to the design intent?
Has the owner's staff been trained to operate the facility?
How many of the systems are operated manually a year after installation?

Positive answers to these and similar questions will ensure that any cost/benefit analysis will demonstrate the value of commissioning.

To ensure that a commissioning process is successful, one must avoid common mistakes. A commissioning plan is a customized approach to ensuring that all the systems operate in the most effective and efficient manner. A poor commissioning plan will deliver poor results. A common mistake is to use an existing commissioning plan and simply insert it into a specification to address commissioning. Each commissioning plan should be specifically tailored to the project to be commissioned.

Also, perhaps due to ill-conceived budget constraints, commissioning is implemented only in the construction phase. Such constraints are ill conceived because the cost of early involvement of the commissioning agent in the design phases is insignificant compared with the cost of correcting design defects in the construction phase. Significant cost savings can arise from identifying design issues prior to construction. Studies have shown that 80% of the cost of commissioning occurs in the construction phase.[2] Also, the later the commissioning process starts, the more confrontational commissioning becomes, making it more expensive to implement later in the process.[2] Therefore, adopting commissioning early in the project is a key success factor.

Value engineering often results in ill-informed, last-minute design changes that have an adverse and unintended impact on the overall building performance and energy use.[3] By ensuring that the commissioning process includes careful evaluation of all value engineering proposals, the commissioning agent and owner can avoid such costly mistakes.

Finally, the commissioning agent's incentive structure should not be tied to the number of issues brought to light during the commissioning process, as this can create an antagonistic environment that may create more problems than it solves. Instead, the incentive structure should be outcome based and the questions outlined above regarding compliance with design intent, training results, and postacceptance performance provide excellent bases for a positive incentive structure.

CONCLUSION

Commissioning should be performed on all but the most simplistic of new construction projects. The benefits of commissioning include:

- Optimization of building performance
  - Enhanced operation of building systems
  - Better-prepared maintenance staff
  - Comprehensive documentation of systems
  - Increased energy efficiency
  - Improved quality of construction
- Reduced facility life-cycle cost
  - Reduced impact of design changes
  - Fewer change orders
  - Fewer project delays
  - Less rework or postconstruction corrective work
  - Reduced energy costs
- Increased occupant satisfaction
  - Shortened turnover transition period
  - Improved system operation
  - Improved system reliability

With these benefits, owners and contractors alike should adopt the commissioning process as the best way to ensure cost-efficient construction and the surest way to a successful construction project.

Coal–Comm

REFERENCES

1. Quantum Consulting, Inc. *Commissioning in Public Buildings: Market Progress Evaluation Report*, June 2005. Available at: www.nwalliance.org/resources/reports/141.pdf (accessed).
2. ACG AABC Commissioning Group. ACG Commissioning Guidelines 2005. Available at: www.commissioning.org/commissioningguideline/ACGCommissioningGuideline.pdf (accessed).
3. Mills, E.; Friedman, H.; Powell, T. et al. *The Cost-Effectiveness of Commercial-Buildings Commissioning*, December 2004. Available at: http://eetd.lbl.gov/Emills/PUBS/Cx-Costs-Benefits.html (accessed).
4. American Society of Heating. *Refrigerating and Air-Conditioning Engineers*, ASHRAE Guidelines 1–1996.
5. Green Building Council Web site. Available at: www.usgbc.org/leed (accessed October 2005).

BIBLIOGRAPHY

1. SBW Consulting, Inc. *Cost-Benefit Analysis for the Commissioning in Public Buildings Project*, May 2004. Available at: www.nwalliance.org/resources/documents/CPBReport.pdf (accessed).
2. Turner, W.C. *Energy Management Handbook*, 5th Ed.; Fairmont Press: Lilburn, GA, 2005.
3. Research News Berkeley Lab. *Berkeley Lab Will Develop Energy-Efficient Building Operation Curriculum for Community Colleges*, December 2004. http://www.lbl.gov/Science-Articles/Archive/EETD-college-curriculum.html.
4. Interview with Richard Holman. *Director of Commissioning and Field Services for Affiliated Engineers, Inc.* Walnut Creek, CA.

Commissioning: Retrocommissioning

Stephany L. Cull
RetroCom Energy Strategies, Inc., Elk Grove, California, U.S.A.

Abstract
This entry examines the practice of commissioning in existing buildings, or retrocommissioning. It provides a definition and a practical understanding of the retrocommissioning process, outlines the energy and nonenergy benefits that result, and examines the link between retrocommissioning and maintenance activities. It also explains the relationship between retrocommissioning and the U.S. Green Building Council's Leadership in Energy and Environmental Design (LEED) certification process and the energy conservation outcomes of retrocommissioning application.

INTRODUCTION

Retrocommissioning is a popular method for reducing energy and operating costs in all types of large, existing buildings. This popularity is due largely to the fact that retrocommissioning pays for itself quickly through the energy and operating cost reductions it produces. Aside from its abundant energy conservation potential, it also offers additional benefits. Depending upon the nature of its application, retrocommissioning can decrease demand maintenance frequency and occupant comfort complaints, improve indoor air quality, and enhance building maintenance staff productivity.

Retrocommissioning is not a one-time event in the life of a building. Rather, its long-term benefits are better realized through a continuous or ongoing approach that is supported by appropriate maintenance activities, throughout the life of a building.

This entry provides a definition of retrocommissioning and how it differs from commissioning and recommissioning. It offers an overview of the retrocommissioning process, the energy and nonenergy benefits, methods used to maintain those benefits, and some typical results of this application. Although a comprehensive retrocommissioning project could extend to all of a building's components, such as operating equipment and systems, core, shell, envelope, and finishes, this discussion focuses on those issues that impact energy utilization, operating costs, and the U.S. Green Building Council's Leadership in Energy and Environmental Design (LEED) certification program.

Keywords: Retrocommissioning; Recommissioning; Commissioning; Continuous Commissioning®; Supported Retro Commissioning℠; Building tune-up; Building maintenance; Maintenance management; Indoor air quality; Energy management; Energy efficiency; Energy conservation; LEED; Green buildings.

DEFINING RETROCOMMISSIONING

Where commissioning refers to a function that occurs during the construction of a building, to ensure that a new construction project meets design intent, and recommissioning is the periodic reapplication of that same function during the life of a previously commissioned facility, retrocommissioning is essentially the commissioning of an existing building that was never commissioned. All three terms refer to a systematic quality management process designed to ensure that building systems and equipment, whether or not they consume energy, function as intended by design, or based upon current use requirements.

Historically, commissioning has not been widely embraced in the new construction industry and its application is currently limited. This is due largely to the lack of hard data concerning the benefits of commissioning and a perception that its application adds unnecessary costs to a construction project. The result is an inventory of buildings that were never commissioned, many of which suffer from undiscovered deficiencies left over from the construction phase. This unfortunate circumstance is becoming more evident to building owners, developers, designers, and contractors and the practice of commissioning, recommissioning, and retrocommissioning is now receiving increased attention.

Retrocommissioning is widely regarded as a more challenging activity than commissioning or recommissioning. This is the case because buildings that were not commissioned during the construction phase generally lack comprehensive or updated reference documentation that is critical to understanding the design and operating intent of building components and systems. Without this up-to-date documentation, the retrocommissioning provider must gather information and perform functional testing "retroactively" to determine the existing performance characteristics for building systems, and then make adjustments to compensate for changes in building configuration and usage requirements. When complete, a retrocommissioning project should provide new and updated documentation sufficient for future recommissioning.

Understandably, and depending upon the age of the building, much could have changed since original construction. This can include changes to the physical configuration of occupied spaces, the addition of equipment and fixtures, new operating sequences, and altered performance characteristics of installed equipment. In large facilities with complex building systems, understanding and documenting these changes can be a time consuming and expensive process. For this reason, many retrocommissioning projects tend to focus only on building systems that have a history of chronic performance problems or those systems that have an impact on energy or operating costs. Projects such as these are often referred to as building tune-ups and they focus more on the effective and efficient operation of a building, rather than static design issues and equipment retrofits. Heating, ventilating and air conditioning equipment, related mechanical and electrical systems, building management and control systems, lighting systems, domestic water systems, and the building envelope are all good candidates for this approach.

THE RETROCOMMISSIONING PROCESS

Retrocommissioning normally includes four phases of work. These phases are used to determine the available opportunities on a broad scale, put a plan in place to address those opportunities, implement that plan, and document results. Although each phase in the retrocommissioning process may require additional components of work depending upon project objectives or complexity of the tasks, most projects follow similar sequences. For the purposes of this discussion, it is generally accepted that a retrocommissioning project will include the following steps:

- A preliminary evaluation of building systems and operating characteristics designed to gather initial data and determine potential project requirements and project scope
- A detailed assessment and project development phase where comprehensive observation, monitoring, testing, and project design takes place
- A project implementation phase where chosen recommendations are implemented
- A documentation and training phase where post-project building characteristics, operating procedures, and equipment conditions are documented, and building operator training is conducted

During the preliminary evaluation stage, the provider will determine the general nature of the project and attempt to come to some preliminary conclusions regarding potential solutions. This phase includes discussions concerning the owner's project requirements, project goals, and the analysis of available data concerning building design and design intent. Current usage requirements, equipment inventories, energy consumption and costs, and occupant complaint histories are examined. Demand maintenance evaluation, equipment performance issues, operating staff interviews, and first hand observation of building system conditions are also carried out. At the conclusion of the preliminary evaluation the retrocommissioning provider should have sufficient information to approximate energy and operating cost reductions the project could produce, and the costs to deliver those reductions. This information is normally summated in a preliminary evaluation report.

The detailed assessment and project development stage of a retrocommissioning project is normally the phase where more definitive conclusions are reached regarding the current operating performance of equipment and systems within the building. At this time, monitoring devices are installed on various pieces of equipment and in occupied spaces, in order to collect and verify current operating data. Information about occupancy patterns, equipment scheduling, control sequences, temperatures, pressures, flows and loads can provide valuable insights into building and systems performance.

In many cases the building will be controlled by a building management system (BMS) or direct digital control (DDC) system capable of collecting data on equipment operation. This data can be extremely useful for populating trend logs that provide real time information about building operation. The degree to which a BMS is capable of collecting this data will determine the need for independent or standalone data collection devices.

The detailed assessment and project development phase is also the time when the condition and efficiency ratings of building equipment is evaluated, and utility cost, consumption and demand profiles are confirmed. At this point equipment demand maintenance histories are reviewed for trends, maintenance and operation procedures are examined, and building operator skills assessed. At the completion of this phase, the provider should have a detailed understanding of the current operating parameters of building systems and equipment, and will have identified the specific strategies and measures to be used in mitigating equipment performance issues and reducing energy and operating costs. The provider will typically provide a detailed assessment and project development report at this stage. This report will usually contain all the relevant information required to make a decision on the financial merits of the project.

During the third phase of a retrocommissioning project the recommended measures that the owner has selected are implemented. This stage of the project is guided by the implementation plan developed during the project development phase. Although it is typical for the provider to manage the quality of the implementation process, individual components of the work are usually performed by third party providers who have specific expertise related to each of the measure requirements. The implementation stage is an excellent time for building maintenance and operations staff to become more familiar with the operational and equipment improvements taking place in their building. Their participation in this phase of the work can be a valuable learning opportunity because it offers a hands-on understanding of the process and a foundation for ongoing commissioning activities in the future. At the conclusion of project implementation, the retrocommissioning provider will have completed the recommended measures outlined in the implementation plan, and will have put in place any project monitoring requirements that were included in that plan.

The final stage of retrocommissioning includes the preparation of as-built documentation concerning design and performance changes made as part of the implementation plan. This should provide the owner with all documentation associated with functional testing and load calculations, along with updates to drawings, specifications, and changes to equipment operation. Information concerning system set-points, operating schedules, operations and maintenance manuals, and any additional changes or alterations made during project implementation should also be provided.

It is also normal at this time to implement any project outcome monitoring requirements and to initiate maintenance and operations training that was identified during the project development phase. Aside from contractual obligations that may require return visits to perform follow-up training or provide project monitoring reports, the retrocommissioning provider's duties are generally complete at this point.

Although the four phases of work described here are widely accepted as standard for core retrocommissioning applications, variations are not uncommon. Unique site conditions, project outcome expectations, the skill level of existing building operations staff, and other factors can all have an impact on sequence and content. In some cases, the relationship past the point of project completion continues because the owner perceives a value in retaining the provider in a project-monitoring capacity. Some providers offer a comprehensive package of services that "support" the owner's operations staff in their efforts to continuously commission their building beyond project completion. Still others will provide a periodic review and reporting function designed to validate projected energy and operational improvements that occurred as a result of the retrocommissioning project. The value and selection of these additional services will always be unique to the circumstances of the project.

RETROCOMMISSIONING AS AN ENERGY CONSERVATION TOOL

Retrocommissioning has the potential to produce large reductions in energy use at a relatively low cost. For this reason, it is becoming a popular cost avoidance tool among building owners and operators. To understand why these energy savings are available it is must be understood that most existing buildings were not commissioned when they were constructed. As a consequence, many buildings do not meet design intent at completion. The result is a large inventory of buildings that are likely to be operating inefficiently. That is not to say that construction projects lack quality control or specification compliance. Simply, it suggests that the historical effectiveness of the quality control mechanism during the construction process is insufficient to fully ensure that a new building operates in accordance with design intent. When factors such as building age, traditionally low investments in maintenance staff training, changes in space configuration, and multiple adjustments to building systems over time are added to that equation, the result is often a building that fails to operate in an energy efficient manner.

Unfortunately, buildings that were not commissioned during construction will have the greatest degree of problems in the very systems that are the most responsible for energy consumption. Heating, ventilating and air conditioning systems; supporting mechanical and electrical equipment; and the controls that govern the operation of these systems play a critical role in the energy profile of any building. In the absence of a comprehensive quality management process during the construction, much can be overlooked. It is not unusual to discover:

- Pumps installed backwards
- Fans and terminal boxes that produce incorrect air volumes
- Missing dampers and damper motors
- Economizers that are incorrectly sequenced
- Disconnected valve actuators
- Systems or equipment that is undersized or oversized
- Inappropriate building management and control strategies
- Simultaneous heating and cooling
- Lighting systems that remain active past occupancy
- Voids in the building envelope
- Equipment or devices that are missing altogether

Although these sort of static deficiencies are not uncommon, they are not the only source for concern. Another, and perhaps more important factor, is the impact that a lack of effective documentation and training at building turnover can have. Recognizing that the cost to operate a building is significantly greater over its lifetime than the cost of original construction, it must be accepted that operations and maintenance plays a critical role in a lifecycle cost analysis. Projects that lack sufficient documentation and allowances for training at completion simply have a greater likelihood for inefficient operation. Unfortunately poor documentation and ineffective training is more the rule than the exception.

Overall, retrocommissioning is a powerful tool for improving energy efficiency and building operator effectiveness. In a 2004 study sponsored by the U.S. Department of Energy[1] that focused on the results of retrocommissioning applications across the country, it was found that, on average, each building had a total of 32 deficiencies. In one building, a total of 640 deficiencies were discovered. Of all the deficiencies found as a result of this study, 85% were related to heating, ventilating, and air conditioning systems. Often, deficiencies like these go undetected for years, causing repeated comfort complaints from building occupants, wasting energy, and increasing the rates of equipment failure and repair costs. Given the frequency of these sorts of deficiencies in new buildings, it should not be surprising that building owners are reaping large energy conservation and operating cost rewards through the application of retrocommissioning.

Although the magnitude of savings that retrocommissioning can produce is dependent on the types of deficiencies discovered, the research sponsored by the U.S. Department of Energy[1] suggests that energy savings can be very significant. In that same 2004 study, which took into account the results from 106 separate projects, energy savings of between 7 and 29% were reported, with paybacks ranging from 0.2 to 2.1 years. Additional data from that study is shown in Fig. 1 below.

NONENERGY BENEFITS OF RETROCOMMISSIONING

In addition to the reasonably quantifiable energy conservation benefits of retrocommissioning, there exists a group of nonenergy benefits that are more difficult to quantify. Although these benefits will have differing values depending upon the type of building and owner motivations, they are none the less important considerations in a decision to proceed with a retrocommissioning project. For example, in the 2004 U.S. Department of Energy[1] study of existing building commissioning, it was discovered that although energy conservation was a primary motivator for 94% of the projects reviewed, a surprising number of projects were motivated by nonenergy benefits. See Fig. 2 below.

As evidenced by this study, issues such as systems performance, improved thermal comfort, and indoor air quality were all deemed to be important factors in choosing to proceed with a retrocommissioning project.

These benefits are difficult to quantify in monetary terms because they are the outcomes of equipment operation or building performance. However, when they are viewed in the converse, their value becomes clearer. For instance, in an office building where the loss of a major tenant can be a costly event to the landlord, maintaining a healthy, comfortable, and productive indoor environment is critical from both a business perspective and a liability standpoint. The value of the benefit in this case would accrue to the avoided costs of finding a new tenant for a vacated space and the avoidance of potential litigation costs associated with an indoor air quality problem.

In certain types of buildings, energy conservation investments can produce other types of financial benefits. An example of this would be the relationship between reductions in operating costs and asset valuation in the commercial real estate industry. According to the ENERGY STAR® Program, every dollar invested in energy efficiency, at a 20%–30% saving rate, is equivalent to increasing net operating income by 3%–4%, and Net Asset Value by $2.50–$3.75. Comparisons like these can be made for all kinds of buildings in all major industries.

An additional nonenergy benefit of retrocommissioning can be derived from improvements in productivity for maintenance and operations staff. This benefit can be sourced from the reductions in occupant comfort complaints that result from improvements in building systems performance. With this reduction comes improvement in productivity or the amount of time that existing staff can expend on deferred maintenance items, other specific activities that impact energy use, or training opportunities that support those functions. Once again, these improvements are difficult to quantify in monetary terms and studies to date have failed to produce definitive findings. It is not difficult to accept, however, that improvements such as these can only have a positive impact on a building operator's time and on the overall cost of operations for any building.

RETROCOMMISSIONING AND MAINTENANCE ACTIVITIES

Retrocommissioning can provide a host of benefits in existing buildings, and studies conducted by many respected individuals and organizations supports that observation. Maintaining those benefits through the life of a building is, however, largely a function of how well the building is operated and maintained during the post-retrocommissioning period. It has been revealed through several studies that the quality of maintenance activities will impact the persistence of energy savings, and equally, the nonenergy benefits that are dependent upon equipment performance.

The relationship between maintenance activities and the persistence of the energy and non-energy benefits of retrocommissioning has received considerable attention. In a 2003 study[2] jointly sponsored by the Department of Energy and the California Energy Commission, the researchers noted that energy savings that averaged 41% of total energy used in a building decreased by 17% over

	Units	Number of projects	Min	Bottom 25%	Median	Average	Top 25%	Max
Commissioned floor area	ft²	106	5,690	95,101	151,000	209,729	271,650	1,014,133
Commissioning Costs								
Total	$2003/building	102	3,214	26,112	33,696	46,442	45,862	476,554
Normalized - excluding non-energy impacts, NEIs*	$2003/ft²	102	0.03	0.13	0.27	0.41	0.45	3.86
Normalized - only for cases including non-energy impacts, NEIs*	$2003/ft²	11	-0.27	0.04	0.17	0.41	0.45	1.88
Cx agent fee as percentage of total commissioning fee	%	9	32%	35%	67%	57%	71%	76%
Costs paid by:								
Building owner	%	31	0%	32%	50%	47%	50%	100%
Utility (e.g. as rebate)	%	48	20%	50%	84%	75%	100%	100%
Other (e.g. research grant)	%	7	33%	100%	100%	90%	100%	100%
Utility rebates (included in above costs)	$2003/building	48	917	11,932	20,500	23,685	25,000	76,725
as % of total costs	%	48	20%	50%	84%	75%	100%	100%
Deficiencies								
Per building	Number/building	85	0.7	5.0	11	32	21.0	640.0
Per 100kft2	Number/100kft²	85	0.1	2.8	6	24	18.3	225.6
Measures								
Per building	Number/building	75	1.0	4.5	9.0	20.3	18.0	481.0
Per 100kft2	Number/100kft²	66	0.1	2.5	5.9	8.6	12.7	218.6
Total Energy Cost Saving								
Raw data (mixed energy prices and years)	nominal $/building-yr	100	-25,752	11,739	33,629	66,489	75,940	879,101
Local energy prices	$2003/building-yr	100	-26,595	13,351	37,376	75,393	80,615	1,034,667
Standardized US-average energy prices	$2003/building-yr	57	-39,043	14,646	44,629	105,156	98,708	1,776,371
Percent energy bill savings	%	74	-3%	7%	15%	18%	28%	54%
Normalized Energy Cost Savings								
Raw data (mixed energy prices and years)	nominal $/ft²-yr	100	-0.09	0.11	0.24	0.42	0.46	3.83
Local energy prices	$2003/ft²-yr	100	-0.09	0.11	0.27	0.47	0.52	4.33
Standardized US-average energy price	$2003/ft²-yr	56	-0.13	0.11	0.26	0.54	0.72	3.23
Monetized non-energy Impacts (one-time)								
Per project	$2003/project (1000s)	10	-281	-31	-17	-45	-11	-1
Normalized by floor area	$2003/ft2-yr	10	-0.55	-0.45	-0.18	-0.26	-0.10	0.00
Energy Savings								
Electricity	kWh/ft²-yr	57	-0.70	0.64	1.7	2.2	2.76	9.72
Percent savings	%	46	-5%	5%	9%	11%	15%	36%
Peak electrical power**	W/ft²	6	0.1	0.4	0.6	0.7	0.8	1.6
Percent savings	%	3	1%	2%	2%	7%	9%	17%
Fuel	kBTU/ft²-yr	29	-14.2	2.3	6.5	15.6	13.5	209.5
Percent savings	%	19	-16%	1%	6%	13%	23%	67%
Thermal (chilled water, hot water, steam)	kBTU/ft²-yr	19	6	32	64	94	122	356
Percent savings	%	16	13%	23%	36%	37%	48%	63%
Total	kBTU/ft²-yr	57	-15	7	17.0	49.3	56	357
Percent savings	%	46	-7%	7%	15%	19%	29%	57%
Payback Times [undiscounted]								
Raw data (mixed energy prices and years)	years	99	-1.5	0.4	1.0	2.1	2.0	20.7
Local energy prices and inflation-corrected cx costs	years	99	-1.5	0.3	1.0	2.1	2.4	26.1
Standardized U.S. energy prices and inflation-corrected cx costs	years	59	-1.0	0.2	0.7	1.7	2.1	10.4

* Non-energy impacts (NEIs) include increases or decreases in first or operating costs due to changes in maintenance costs, contractor callbacks, equipment life, and

** Most are averaged over the entire year, hence true "peak" savings are significantly higher than shown here.

Fig. 1 Result summary with quartile analysis—existing buildings.
Source: Reprinted from "The Cost Effectiveness of Commercial Buildings Commissioning" LBNL Publication No. 56637 (see Ref. 1).

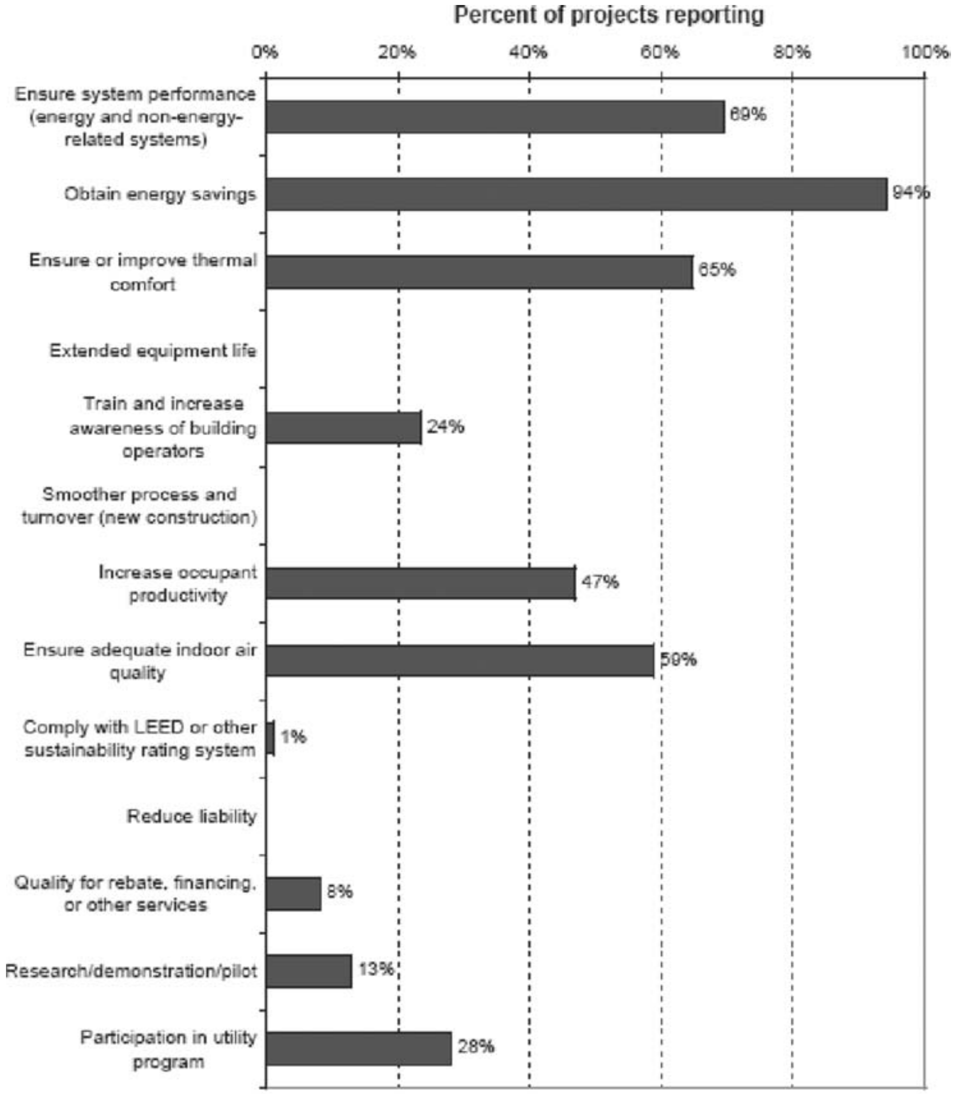


Fig. 2 Reasons for existing building commissioning.
Source: Reprinted from "The Cost Effectiveness of Commercial Buildings Commissioning" LBNL Publication No. 56637 (see Ref. 1).

two years. These findings also indicated that the long-term persistence of energy savings hinged on the abilities of the building operators to troubleshoot and understand how the systems in the building were supposed to operate.

In concluding why retrocommissioning benefits would persist in some applications and not in others, the study observed that persistence was influenced by a group of factors. The most notable of these appeared to be the working environment and operator training. Successful projects, where savings tended to persist, were those with working environments that provided high quality operator training, time to study and optimize building operation, and a management group that was focused on optimizing the performance of the building and reducing energy costs. In addition to the working environment and training, the study concluded that performance tracking and adequate documentation could impact benefit persistence as well. In the case of performance tracking it was suggested that energy use tracking and trend data analysis were important factors for persistence. Proper documentation concerning building equipment and its operating intent could also provide building operators with information on how to effectively operate the building.

Creating benefit persistence through building operator training can have a profound affect on how valuable a retrocommissioning project can be through the life of any building. This factor has been particularly well documented through the experiences at Texas A&M University over a number of years. Since 1993 the school has deployed a process called Continuous Commissioning® in more than 130 large campus buildings and has made training for building operators in this process a cornerstone of its maintenance efforts. The results of this focus have been dramatic, producing maintained energy savings for the school of between 15 and 25%. In a 2001 study conducted under the California Energy Commission's

Public Interest Energy Research Program and involving ten buildings at Texas A&M, the Continuous Commissioning process was projected to deliver $4,255,000 in energy savings over a four-year projection.[3]

Given the magnitude of energy savings and other benefits available from retrocommissioning, and the impact that maintenance functions and training have on the persistence of those savings, it is not difficult to conclude that a well-trained, effective, and supported maintenance program is essential for any building.

RETROCOMMISSIONING AND LEED CERTIFICATION

Commissioning and retrocommissioning have both found their way into the U.S. Green Building Council's LEED certification and point award process. This clearly illustrates the value that the USGBC places on these processes and provides further evidence that their application can have a positive impact on environmental and energy issues for any building.

The LEED certification process provides award points that encourage "green" building design, construction, and operation. Depending upon award points attained, a building can be ranked as LEED Certified, LEED Silver, LEED Gold, or LEED Platinum. It is important to point out that these award rankings can be achieved only if a new construction project is commissioned or an existing building is retrocommissioned. It is also important to note that in all rankings, LEED places a great deal of emphasis on the very systems in a building that account for the greatest impact on energy costs. This emphasis is similar for commissioning and retrocommissioning, and in many ways these processes, and the intent of the LEED rating system, are functionally complementary. Both have a positive affect on energy consumption, and therefore, a similar impact on the environmental footprint a building will make.

In existing buildings, retrocommissioning is regarded as a cornerstone of the LEED certification process. The LEED points available for existing buildings focus attention on building operations, testing, monitoring, repeated commissioning, and continuous improvement. Points are available for such things as the development of a Building Operations Plan, a commitment to 24 hours of training for building operators each year, performance monitoring of equipment operation and maintenance best practices. It is not coincidental that the most successful retrocommissioning projects encourage and support similar activities.

CONCLUSIONS

Retrocommissioning offers excellent energy conservation and operational improvement opportunities for existing buildings, with minimal investment. It can improve the effectiveness of operations and maintenance staff, provide a group of nonenergy benefits and can assist in the attainment of a LEED certification. However, the most important factor the reader can take away from this section is that commissioning during construction can eliminate many of the problems that owners inherit at the acceptance of their project. Ultimately, as commissioning is more widely adopted in new construction projects, the need for retrocommissioning could be eliminated altogether.

In the meantime, however, retrocommissioning is gaining an impressive following of supporters.

It is encouraging that the Department of Energy, ENERGY STAR®, American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), the Association of Energy Engineers, the California Department of General Services, the California Energy Commission, The U.S. Green Building Council, and many others now recognize the impact that this application brings to energy conservation and environmental protection. In California, retrocommissioning has found its way into the State's Green Building Action Plan by requiring that all buildings 50,000 ft^2 and over be retrocommissioned and that periodic recommissioning take place in the following years. In another signal of the value of this application, utilities in California are now offering rebates and incentives to encourage building operators and owners to implement retrocommissioning in their buildings.

Given the numbers of buildings in North America that have not been commissioned, and the ever increasing pressures to reduce energy consumption, it is not difficult to envision that retrocommissioning will become a very popular energy conservation measure in the years to come.

REFERENCES

1. Mills, E.; Friedman, H.; Powell, T.; et al. *The Cost Effectiveness of Commercial Buildings Commissioning-A Meta Analysis of Energy and Non-Energy Impacts in Existing Buildings and New Construction in the United States*; Lawrence Berkeley National Laboratory, Portland Energy Conservation Inc., Energy Systems Laboratory-Texas A&M University, Building Technologies Program, U.S. Department of Energy, 2004, Lawrence Berkeley National Laboratory Report No. 56637, 31,34,36, http://eetd.lbl.gov/emills/PUBS/Cx-Costs-Benefits.html.
2. Friedman, H.; Potter, A.; Haasl, T.; Claridge, D. *Strategies for Improving Persistence of Commissioning Benefits*; Portland Energy Conservation Inc., Texas A&M University Energy Systems Laboratory, U.S. Department of Energy, California Energy Commission, 2003; 2,4.
3. Turner, D.; Claridge, D.; Deng, S.; et al. *Persistence of Savings Obtained from Continuous Commissioning*; Texas A&M University Energy Systems Laboratory, Texas A&M University Physical Plant, University of Nebraska, Omaha, California Energy Commission, 2001; 13, HPCBS No. E5P2.2T5a2.

Compressed Air Control Systems

Bill Allemon
North American Energy Efficiency, Ford Land, Dearborn, Michigan, U.S.A.

Rick Avery
Sam Prudhomme
Bay Controls, LLC, Maumee, Ohio, U.S.A.

Abstract
This Compressor Control Systems entry provides an outline of the strategies used to manage compressed air generation at a typical manufacturing facility and discusses the components of a compressed air system, typical control methodology, and energy saving strategies. Compressed air distribution, metering, and monitoring is also reviewed. This entry is meant to give the reader a basic understanding of compressed air generation and management fundamentals.

INTRODUCTION

Compressor systems are benefiting from increasingly advanced control technologies to realize improved efficiency, safety, and operational benefits. Compressor controls now use microprocessors, computer networking, sophisticated control algorithms, and Web-based monitoring to provide superior control capabilities and features.

Today's control systems do more than just operate the air compressor. Using sophisticated technologies, advanced controls can significantly reduce energy expense and lower maintenance costs. Automation is another important feature, allowing much of the standard compressor system operation to be managed by software. By monitoring and controlling compressor auxiliary equipment, modern controls can help ensure that high-quality air at the lowest cost is reaching the end user. Zone management provides the ability to regulate air use in various plant departments. Remote and Web-based integration systems improve and streamline system control and management. Extensive data gathering and reporting provides real-time information to help drive business decisions, such as evaluating system expansion, identifying additional savings opportunities, and managing predictive maintenance.

Modern compressor control systems address the entire compressed air infrastructure. Today, a complete control approach can include the following facets of a compressor system: air production controls for air compressors and the motors that power them; air quality monitoring for dryers and other air conditioning equipment; distribution control in the form of zone management; and integration, management, and metering products to improve operations and decision making.

Keywords: Compressor; Control; Air; Dryer; Centrifugal; Pressure; Surge; Blowoff; Leak; Automation; Regulate; System; Efficiency.

COMPRESSOR CONTROL SYSTEM COMPONENTS

Air Production Control

The supply side of a compressor system consists primarily of the compressor and the motor. Control systems have several important tasks: to ensure that the compressor system produces enough air to meet plant demand; to keep the compressor and motor running without costly shutdowns; to operate the system as efficiently as possible; and to prevent damage to the major pieces of equipment.

Older compressors use pneumatic and electro-mechanical control systems. Modern microprocessor-based controls improve performance by offering precise pressure regulation, networked capacity control, and additional features. For centrifugal compressors, advanced surge control can dramatically increase efficiency by reducing wasted air blowoff.

Electronic controls are available for all makes and models of compressors. To provide maximum efficiency and automation, a control system should electronically network the compressors, regardless of type or make. The primary goal of a control system is to ensure that the compressors produce enough air to meet plant demand.

Motor Control

Compressor systems begin with the prime mover, often an electric motor. A well-functioning motor is obviously crucial to the operation of a compressed air system.

Effective electric motor control prevents damage and helps alleviate common electric power concerns.

Primary Motor Control

Modern electric motor controls offer a number of features that improve motor reliability. Extensive motor protection is built into these products, preventing motor overload and burnout that can result from drawing too much power. User-friendly, man–machine interfaces, with Liquid Crystal Display (LCD) screens and micro keypads, are now common features. These allow for easy initial configuration and modification of the control's operating parameters. Some motor controls are based on solid state components, which require less maintenance and replacement of costly electrical parts that otherwise wear out over time.

An important element of modern motor controls is the soft start function, wherein the motor is gradually brought up to its maximum power. Compared to a typical high torque start, a slow start greatly reduces mechanical and electrical system shock to the motor and attached compressor. This results in less wear and tear from starts and stops and reduces long-term maintenance of expensive motors. A soft start also reduces the initial electrical current inrush, placing less stress on the facility's electrical system and alleviating related low power problems.

Ride Through Motor Control

An important new auxiliary motor control is the ride through controller. This device addresses the problem of momentary interruptions to the motor's power supply, which, no matter how short, can often cause motor shutdown. Motor shutdown stops the supply of compressed air, halts the facility's production, and results in lost productivity and increased costs. A ride through device has the ability to keep the motor operating during momentary power interruptions and voltage sags. Because these are lower cost auxiliary controllers, they can often quickly justify installation costs.

Compressor Control

The most fundamental part of a compressor control system is the controller on the individual compressor. These devices are designed to protect the compressor from damage, operate it as efficiently as possible, and automate recurring control actions. Modern compressor controls offer features far beyond those of their predecessors; powerful microprocessor-based controls are becoming increasingly popular as their significant advantages become more widely known.

Compressor Control Types

Pneumatic

Many existing air compressors utilize electro-pneumatic control systems. These systems use electric and mechanically activated devices such as pressure switches, solenoid valves, and metering pins. To address monitoring and compressor protection, electro-pneumatic systems typically feature a series of mechanical trip switches which shut down the compressor when pressures or temperatures reach critical levels. With their mechanically limited control logic, pneumatic controls offer only a basic set of operating functions—usually simple compressor modulation and monitoring. In addition, the response time of pneumatic controls is typically inferior to other types, reducing the effectiveness and efficiency of control actions.

Programmable Logic Controllers

Programmable logic controller (PLC) systems are in common use today, and they represent a significant improvement over pneumatic-based systems. Programmable logic controller control systems utilize digital and analog control and monitoring instruments. The responsiveness and accuracy of these devices enable greater compressor efficiency. Additionally, because PLCs are electronically based, they can offer more functions, such as sequencing and advanced control and monitoring. Fundamental speed and performance limitations of PLC hardware, however, still leave room for improvement.

Microprocessor

The market for microprocessor-based controls is growing rapidly due to their comprehensive features and superior performance. The control and monitoring accuracy of these systems is excellent, allowing for tight pressure regulation and advanced protection strategies. Because these controls are based on hardware similar to modern computers, they can offer additional features such as sophisticated networking, multiple compressor control, and extensive operational record keeping. Many of the control benefits and features listed in this document are best implemented by these types of control systems.

Protection and Safety

On the most basic level, compressor controls must keep the machine running safely and reliably, preventing both serious compressor failure and more common mechanical damage.

Plant Safety

Maintaining plant safety by preventing catastrophic compressor failure is one of the most fundamental responsibilities of a compressor control. One example of severe compressor failure is when incompressible water builds up in a reciprocating compressor's compression chambers; if enough accumulates, the compression action can cause the equipment to physically fail, sometimes explosively. Other safety concerns include overheated and potentially combustible oil entering the air distribution system or violent vibration caused by a severe mechanical failure. Most modern control systems have the ability to shut down the compressor if a serious malfunction is detected.

Machine Protection

On a less extreme scale, compressor controls should guard the compressor from common wear and damage. Advanced control systems use analog and digital monitoring instruments to check the compressor for abnormal operating conditions. When the control system detects an unsafe measurement, it can either trigger a warning alarm or shut down the unit, depending on the nature of the situation.

Modern control systems examine every relevant operational value of a compressor and continuously check these values against standard ranges. Today's systems can visually display a list of the current values and alarm parameters of every monitoring point, allowing an operator to know precisely what the compressor is doing at any moment. Another useful feature is a recorded history of past operation and alarm events. If a problem occurs, such a record can provide critical diagnostic information to help resolve the issue quickly.

Centrifugal Compressor Surge

On centrifugal compressors, a key control feature is the ability to reduce or eliminate surge. Surge is a phenomenon where compressed air rapidly oscillates backwards then forwards through the compressor; unabated, this can cause severe damage. From a maintenance and protection standpoint, modern controllers offer far superior surge prevention compared to older systems. Advanced systems, with mathematical models of the compressor's surge line and fast control responses, work to keep the compressor out of surge. If surge does occur, these systems can detect the event and adapt the controller's operation to avoid any recurrence.

ENERGY SAVINGS WITH MODERN CONTROLS

One of the primary benefits of an advanced compressor controller is reduced energy expenses. Energy savings can result from any or all of these factors:

- Precise pressure regulation reduces the average system pressure output.
- Networked capacity control coordinates production among multiple compressors for maximum efficiency.
- Advanced centrifugal control can reduce wasted air from blowoff.
- Leak loss reduction is a byproduct of a lower average system pressure.
- Automated load scheduling can shut down or offload compressors when plant demand is lower.
- Proper intercooler control ensures better compressor efficiency.

Precise Pressure Regulation

Significant energy savings can be realized by lowering compressor discharge output pressure. Older systems are slow and inaccurate, resulting in large plant pressure swings. In order to keep pressure from swinging below the minimum, these older controls commonly maintain an average system pressure much higher than necessary. Modern controllers, with faster, more precise abilities, and sophisticated control strategies, can greatly reduce pressure swings. The smaller pressure range, typically within 2 psi of target pressure, allows a subsequent drop in average pressure setpoint. A 2-psi reduction in system pressure results in an approximately 1% drop in energy use; thus, the potential energy savings are substantial.

Networked Capacity Control

In plants with multiple compressors supplying a common system, uncoordinated compressor operations often offer opportunities for energy savings. Compressors that do not communicate with each other act independently, raising or lowering output as they detect changes in plant demand. This can often result in competing compressors; one compressor may be lowering its output while another is increasing its capacity. Unstable plant pressure levels are one result of this competition. Additionally, the compressors operate at inefficient part load capacity levels.

Sequencers, which were the first solution to this problem, assign compressors different fixed pressure levels at which point they come on or off line. Although this method does prevent competition between the individual compressors, it causes plant pressure to fluctuate along the range of assigned pressure levels (a four-compressor system with pressure intervals of 5 psi will have a 20 psi operating window) and tends to maintain an average system pressure higher than needed.

Advanced control systems with networked capacity control capabilities produce an efficient compressor system operation. Through network communication, the controls automatically operate as many individual compressors at their most efficient (full load) capacity levels as possible. Instead of multiple compressors operating at part load capacities, a single compressor is modulated to meet plant demand. The whole system is coordinated to maintain a single pressure setpoint, providing precise pressure regulation. Should demand fall, compressors are automatically shut down or unloaded, further saving energy. Rising demand will cause another compressor to come online, ensuring stable plant pressure. With such a system, multiple compressor installations can maintain plant pressure while operating individual compressors as efficiently as possible.

System Controllers

Built-in networked capacity control is limited to the more advanced control systems. When installation of these systems is not feasible, similar capabilities can be achieved with a system controller. In order to achieve the benefits of networked control with less capable, mismatched, or incompatible control devices, a system controller is sometimes used to coordinate the individual compressors. These master controllers offer many of the features of networked control, and can operate several compressors at a common plant pressure setpoint.

Surge Control and Blowoff Reduction

For centrifugal compressors, blowoff at minimum capacity is a significant energy waste. Modern controllers can reduce this waste using several methods. Advanced antisurge algorithms provide greater turndown, allowing a larger modulation range before blowoff begins. Additionally, when the minimum capacity point has been reached, the controller can switch the compressor into an unloaded state, where it produces little air and thus blows off little air. In a networked capacity control system, other compressors can modulate output, allowing the centrifugal compressor to operate at higher and more efficient capacity levels. These combined features offer great energy savings potential.

Leak Loss Reduction

Reducing the average pressure setpoint also reduces the amount of air that escapes from existing leaks. Leakage easily can be the largest energy problem in a compressed air system, ranging from 2 to 50% of compressor system capacity. An average plant has a leak rate of about 20% of total air production.

Lower pressure air has less air mass in the same volume. Because the volume of leaking air remains constant at a given pressure, a lower pressure results in less air escaping from existing leaks. Precise pressure control is one way to lower average system pressure.

Load Scheduling

Load scheduling automatically matches the compressor's output pressure with predetermined plant demand. During breaks and off-production periods, a significantly lower pressure often can be maintained in the plant. Any pressure reduction will save notable amounts of energy.

Compressor Intercooler Control

Effective intercooler control will help a compressor operate at top efficiency. Air that has not been cooled adequately by the intercooler will enter the next stage with a larger volume, reducing total compressor output. Alternately, air that is cooled too much can form liquid condensate, which can damage compressor components and increase maintenance costs. Modern control systems usually include intercooler control as a standard feature.

Automation

Automation is a key element of advanced control systems. Automated machine protection, data collection, and start/stop, combined with capacity regulation and load scheduling capabilities, give modern control systems the ability to automate nearly every operation of a compressor system. Remote monitor and control capabilities further reduce dependence on at-the-controller compressor supervision.

Start/Stop

Modern control systems can automate a compressor's start and stop procedures. With monitor and control connections to essential compressor subsystems (motor, lubricators, coolant, etc.), the controller can start and stop the complete compressor station while ensuring safe operation.

In a networked control system, automatic start capability can add additional system reliability. When one compressor is shut down because of a problem, the system can automatically compensate for reduced air supply by bringing additional compressor capacity online. Thus, plant pressure is maintained and production continues with limited interruption.

Scheduling

When start/stop capabilities are combined with a schedule, much of the day-to-day compressor operation can be automated with controls. Once an effective schedule is established, the system can essentially run on autopilot, with little need for immediate operator adjustments.

AIR QUALITY

An often overlooked element of compressor system controls is the equipment used to condition air, which primarily includes dryers, aftercoolers, and filters. These pieces of equipment have the essential task of ensuring that hot, wet air leaving the compressor is converted to high-quality, cool, dry air for use in the facility. The two largest air conditioning problems are insufficient drying and excessive differential pressure drops.

Wet air that enters the distribution system can eventually cause rust formation, leading to clogs and extra wear on end use equipment. If the problem becomes severe enough, it can require a shutdown of portions of the compressor system while corroded piping and failed components are replaced. These humidity problems can be caused by inadequate drying or poor aftercooling.

An improperly sized, poorly maintained, or outdated piece of equipment can cause an excessive pressure drop as air passes through it. Pressure drop across dryers alone can be greater than 6 psi; because a 2-psi pressure change roughly equates to 1% of energy used to compress the air, the opportunities for energy savings are substantial. Excessive pressure drop is a factor that can affect dryers, aftercoolers, and filters.

Monitoring Air Quality

Given that high quality air is so important, there are opportunities for mitigating these issues with monitoring techniques. By monitoring the pressure and quality of the air, both before and after conditioning has taken place, a facility can identify existing and potential problems. With continued monitoring, the results of equipment upgrades or maintenance actions can be verified for effectiveness. This approach is also able to flag when maintenance is necessary to upkeep the quality of air, reduce unnecessary pressure drops, and maintain system efficiency. Monitoring capabilities of this type are most often integrated into an overall compressor management system, where the metered values are displayed on a networked compressor management system workstation.

AIR DISTRIBUTION

Zone Management

An emerging control strategy for the distribution side of a compressed air system is known as zone management, which provides additional opportunities for energy savings and operational improvements.

Zone management technology gives the capability to monitor and control the demand side of a compressed air system. This strategy involves separating an air system into different distribution zones and regulating and metering the air supply to each one. The control of each zone can be scheduled for automatic operation, making complex zone management relatively easy. Zone management opens up new system operation options, such as running zones at different levels of pressure and shutting off zones when they are not in use.

With the ability to monitor the air consumption of each zone, users can gain much greater insight into the operational dynamics of a compressed air system. This advanced metering ability provides an accurate determination of the compressed air energy costs of different plant operations and can provide incentives and justifications for initiatives to reduce those costs.

On a design level, zone management involves logically separating the operations that use compressed air into different zones based upon factors such as concurrent air use, pressure setpoints, and air quality. The air supply to each zone is then individually controlled, allowing the air flow to each zone to be modulated according to the current use in that zone.

With a scheduling function, the system can be configured to automatically raise and lower pressures or to turn the air supply completely off to each individual zone.

Implementing zone management requires a combination of metering and control instruments for each zone, and a master control device to provide the monitoring, control, and scheduling functions. This master control functionality is often an add-on capability of a compressor management system workstation.

Benefits of Zone Management

A facility can realize one or more benefits from a zone management system: reduced compressed air energy consumption, zone cost regulation, increased data collection, air quality monitoring, and zone air leakage measurement.

Reduced Air Use

Zone management often results in a reduction in air demand due to its ability to lower the pressure or completely stop the flow of air to zones that are not in use. Less air is then used to maintain pressure in nonproduction areas and total demand is lowered. Less demand means less compressed air production, which directly reduces energy costs.

Regulate Air Costs

The ability to establish a separate cost center for each air use zone is another important benefit that comes from the capability of monitoring and metering the air use of each zone. With a comprehensive metering program, the facility has the ability to regulate the costs of compressed air for different segments of their production operations. Better information regarding air use and the costs of that use allows for better management decisions to be made.

End to End System Information

When zone management is paired with modern controls for other sections of the compressor system, a facility can have complete end to end system information and management. This system-wide approach enables better understanding of the operational dynamics of the entire compressed air system.

Monitor Air Quality

Air quality can be just as important as energy management. Correct pressure and humidity levels ensure the proper functioning of end use equipment and lower the lifelong maintenance costs of equipment based on wear and tear. Zone management allows constant monitoring of the pressure and dew point of the air entering each zone, ensuring that high-quality air reaches the end use point.

Measure Zone Leakage

Leakage is a frustrating aspect of compressed air systems and one that is difficult to measure. Zone management makes it easier to determine how much air is used and thus wasted by a particular zone when it is not in use. A majority of this wasted air comes from air leaks. By measuring the amount of air leaking from a zone, the facility can make effective decisions regarding leak control programs and the potential return from such efforts.

AIR MANAGEMENT, MONITORING, AND METERING

The final control layer for a compressed air system is an integration strategy that pulls all of the different control and monitoring features into one centralized location. This is usually accomplished through a compressor network, which is then routed to a human machine interface (HMI) workstation. This workstation can be a local monitor and control computer, or a Web-based management and analysis client server.

Facility Intranet Monitoring and Control

When connected to a network of individual compressor controllers, a remote computer station allows for monitor and control of an entire compressed air system from a single convenient location. With access to the entire range of monitored compressor data, the remote station can display the condition of the complete system at any given moment. A remote operator with security access to control capabilities can start, stop, and change the capacity of any compressor from the remote station.

Remote operation also allows an operator to pinpoint compressor problems as they occur. More immediate and precise information directs corrective actions, reducing compressor down time and associated costs from loss of adequate plant air supply.

A workstation connected to a network of compressor controls can also record and store data for the entire compressor system. This wealth of information is then available for analysis, which can quantify compressor system performance over time. Individual compressor problems can be identified and corrected before they become serious. This preventive maintenance reduces operating costs resulting from both inefficient operation and lost production due to compressed air equipment failure.

Below are some of the specific features of a typical HMI compressor workstation:

- A total system view, which displays the operating conditions of the complete compressed air system, including pressure, motor status, and alarm status.
- Complete remote control over every compressor on the network, including starting and stopping, capacity changes, and pressure setpoint changes.
- An individual compressor view, which displays the current readings of the major monitoring points, while graphically showing the user where the points are located on the unit.
- A complete list of all monitoring points for the individual compressor, with alarm and trip values and current status.
- Recorded history of all alarm and trip readings that occur.
- Recorded history of all operator events for the compressor.
- An ability to generate analytical performance reports based on the collected data. These reports can track compressor performance over time and spot trends before they become problems.

Web-Based Management Systems

A Web-based compressor management system offers many similar features to a local compressor network.

One of the central capabilities of a Web system, however, is the ability to provide access to multiple facilities from a standard internet connection.

Usually, a gateway device of some kind is used to connect the facility compressor network to a dedicated communication connection, such as a phone line or broadband connection. An off-site server then collects data from the system and a Web portal provides access to the information. From the site, a user can access real-time monitoring of plant air compressor systems, as well as the operating parameters of each individual compressor. For security purposes, a Web system often does not allow control actions; this prevents unauthorized tampering with the facility's compressor operation. If a company operates multiple plants, the Web site can provide centralized monitoring and metering capabilities for the entire enterprise. The Web-based nature also means that a user can access compressor system information from anywhere there is an internet connection. Additionally, Web systems frequently incorporate built-in efficiency and operations analysis and reporting, giving management powerful tools to measure the performance of their air compressor systems.

Benefits from Whole System Integration

These centralized management, monitoring, and metering systems offer many benefits to the operation of a compressed air system. Some of the more common benefits are described in the following sections.

Internet-Based Access

A Web-based system is accessible from any standard internet connection. This provides operators with monitoring capabilities from office, home, or even Web-enabled phones or PDAs (personal digital assistants, such as Treo or Blackberry).

Cost Metering

An integrated management solution offers the ability to meter the costs of compressed air operations. Comprehensive whole system metering provides operators and management with improved understanding of the costs and operational dynamics of the compressed air system. Real-world data can then be used to identify opportunities and drive business and management decisions.

System Efficiency

The ability of a management system to accurately and consistently calculate the cost of compressed air for single or multiple facilities makes it easier to measure any changes in efficiency and compressed air production. As future investments are made in the compressed air system, a management and metering system can compare system efficiencies before and after improvements are made to validate savings. Using centralized monitoring and analysis capabilities, a management system can often help identify further efficiency gains from operational changes. For example, these systems can help find the most optimal, most efficient mix of compressors to use during different shifts at a facility. Once the optimal mix is found, automatic scheduling ensures that the most energy efficient solution is used.

Measure Air Leakage

Leakage is a notoriously hard quantity to measure in normal circumstances. A system management tool can easily measure the air flow rate of a facility during nonproduction hours; this flow will be a close approximation of a system's air leakage.

Preventive Service Monitoring

The constant monitoring performed can be helpful in identifying and resolving problems with compressors before they become serious. Problems with sensors or other issues can be identified during periodic system reviews and resolved before the issue becomes serious.

Real-Time, Data Driven Troubleshooting

When a problem with a compressor does occur, the data recording and real-time monitoring capabilities of a whole system management solution can provide faster resolution to the problem. With a Web-based solution, experts from outside facility locations can view the current system status and help with troubleshooting.

Comprehensive Overview

Companies with multiple facilities find a Web system very useful in providing a centralized overview of all compressor networks. These tools provide an easy and accessible method to see the current status of every compressor on the system.

CONCLUSION

Modern compressor control systems now offer increased levels of sophistication and opportunities for control of the entire compressor system. A comprehensive, system-wide solution cannot only control and protect individual compressors, but it can also provide powerful new operational and management tools. This new control strategy provides powerful, efficient, and effective control of compressed air systems.

Compressed Air Energy Storage (CAES)☆

David E. Perkins
Active Power, Inc., Austin, Texas, U.S.A.

Abstract
Compressed-air energy storage (CAES) is currently being deployed as an alternative to lead-acid batteries for uninterruptible power supplies. These systems use compressed air supplied from either transport cylinders delivered by local gas services, or from stationary cylinders refilled from on-site compressors to drive a variety of economical expansion engine topologies. Several factors make these systems feasible for use in small-scale CAES systems for load leveling in conjunction with wind or solar energy generation while opportunities exist for improving cycle efficiency.

INTRODUCTION

Large-scale CAES has been successfully used as a means of peak shaving as an alternative to peaking gas turbines. Two such systems are the 290-MW unit in Huntorf, Germany[1] and the 110-MW unit in McIntosh, Alabama.[2] These systems allow independent operation of the compression and expansion processes commonly found in conventional gas turbines. These systems have been called hybrid CAES; in that they continue to use fuel in the expansion process, the benefit being that the turbine need not produce power to drive the compressor when operating from the compressed air reserve. A disadvantage of these systems is that they require the use of fuel, which results in CO_2 emissions, and the heat of compression is discarded thus compromising the cycle efficiency.

One solution to these issues is to use adiabatic compression and thermal energy storage (TES) in place of the combustion process associated with hybrid CAES. Systems employing adiabatic compression and energy storage were explored in the 1970s and have received renewed interest. One program titled Advanced Adiabatic Compressed Air Energy Storage (AA-CAES)[3] is underway in Europe, which is focused on zero-emission storage technology for centralized storage as well as modular products for distributed storage.

Several companies are now offering CAES systems in various configurations as environmentally friendly alternatives to the lead-acid batteries found in uninterruptible power supply (UPS) systems. Earlier this year, active power-introduced products for the UPS market based on its Thermal and Compressed Air Storage (TACAS) technology.[4] Several companies, including a major supplier of photovoltaic cells have expressed interest in the use of this technology for electricity storage generated by renewable sources. At this time, however, systems based on TACAS technology achieve cycle efficiencies between 10 and 15%, since it currently relies on oil-lubricated reciprocating compressors, which are nearly isothermal. This paper presents the current embodiment of the TACAS technology for UPS and discusses proposed enhancements to improve cycle efficiency for electricity storage.

THERMAL AND COMPRESSED AIR STORAGE FOR UPS

A CAES system with thermal energy supplied by the grid has been developed for UPS applications. The output of this system is 85 kW and is capable of delivering power for up to 15-min (21 kWh). A schematic of the system architecture is shown in Fig. 1.

The TACAS technology system uses high-pressure gas cylinders for air storage since volumetric energy density should be compatible with batteries for UPS applications and the use of caverns is impractical. The use of high temperature TES heated by the grid allows higher turbine inlet temperature than would, otherwise, be available from direct expansion of the air from cylinders.

The specific energy available from compressed air is a function of the turbine pressure ratio and inlet temperature. System sizing then becomes a tradeoff between the mass of gas stored in the cylinders and the mass of TES required to heat the gas being delivered to the turbine to achieve a desired inlet temperature. Cost optimization was performed to balance the size of the cylinder banks and TES as a function of the turbine pressure ratio, inlet temperature, and discharge temperature. The turbine pressure ratio was constrained by manufacturing capabilities and the discharge temperature was constrained to be

☆This entry was originally presented as "Compressed-Air Energy Storage for Renewable Energy Sources" at the World Energy Engineering Conference (WEEC), 13–15 September 2006, Washington DC, U.S.A. Reprinted with permission from AEE/Fairmont Press.

Keywords: Energy storage; CAES; Uninterruptible power supplies; UPS; Air turbine generators; Thermal energy storage; Renewable energy.

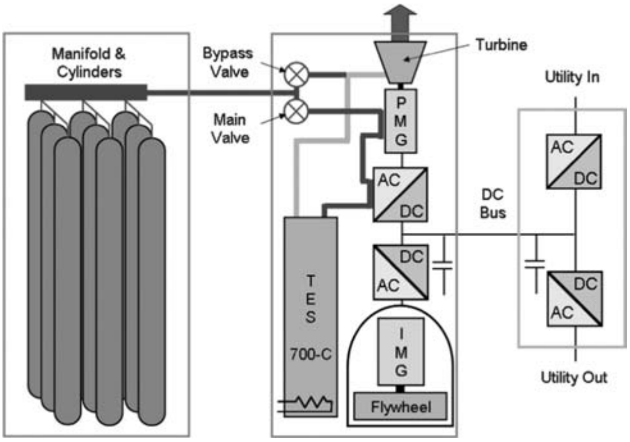


Fig. 1 Thermal and compressed air storage (TACAS) technology for UPS.

suitable for discharge into an interior building space. Maximum storage pressure was constrained by the most economical cylinders and compressors as well as diminishing returns due to significant effects of gas compressibility effects (>10%) above 310 bar [4500 psi].

Small-scale electric power and storage systems are generally more expensive to produce per kW and/or kWh output, so the optimization described above is essential. Furthermore, in order to meet system-cost targets, the simplest designs must be adopted.

EXPANSION TURBINE AND CONTROL

For small-scale systems, single-stage axial-flow impulse turbines offer an efficient and cost effective solution for high-pressure ratio expansion applications, particularly when direct-coupled to a high-speed permanent-magnet generator (PMG). A design speed of 70,000 rpm minimizes the PMG size helping to reduce the mass of expensive permanent magnet material and is a compromise between several machine design factors.

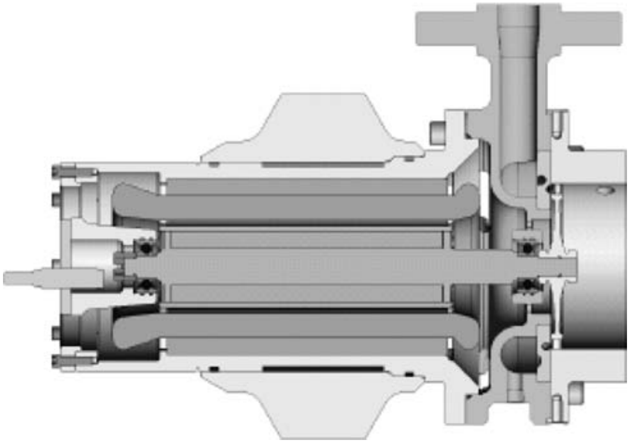


Fig. 2 100-kW single-stage axial-flow air turbine generator.

The system-optimized turbine shown in Fig. 2 operates at 385-psig and 230°C inlet temperature and delivers 100-kW shaft power at an efficiency of approximately 72% (small turbines typically have efficiencies between 70 and 80% whereas larger designs can approach 90%). Nearly the entire temperature drop occurs during expansion across the supersonic converging-diverging nozzles, so the temperature reaching the bladed disk approaches ambient. This allows the use of inexpensive rotor materials with integrally machined blades.

The single-stage turbine readily adapts to overhung-rotor architecture. The short overhang distance and short overall rotor length considerably simplify rotor-dynamic issues. In this particular configuration, rolling-element bearings with elastomer dampers are employed since the operational duty cycle for UPS is low. The design is easily configurable to foil-type gas bearings for higher duty service. Blade reaction is close to zero so axial bearing loads are very low and easily managed. Cooling of the bearing closest to the turbine inlet plenum is a challenge, but effective routing of air released from the control regulators helps to mitigate this. In fact, cooling of the turbine-PMG and its power converter is also provided by the process airflow.

Because the system is presently designed for UPS, transient response is of paramount importance. It takes about 1 s for the turbine-generator to reach full speed while carrying increasing load during the acceleration event. A small flywheel is used to "bridge" the gap between utility outage and full-load turbine output. This flywheel is configured with a bi-directional converter, which allows the flywheel to absorb step unloads and eliminates the need for unloading valves. During discharge, the flywheel also manages small power fluctuations so that turbine speed is held constant.

Simple turbines based on fixed nozzle geometry achieve power regulation through inlet pressure and temperature control. For this system, turbine inlet pressure and temperature are controlled by a pair of dome-loaded pressure regulators using a unique control scheme,[4,5] which routes air through or around the TES. Referring to Fig. 2, Main Regulator 340 controls airflow going to the Heat Exchanging System 350 (in this case, the TES). Bypass Regulator 320 controls airflow around the TES. Orifice 330 in the bypass path provides control stability. By combining flows through regulators 320 and 340 in varying proportions, constant fluid discharge temperature and pressure can be achieved throughout the sliding temperature range of the TES.

Regulators 320 and 340 identified in Fig. 3 are called dome-loaded regulators. This type of regulator is often employed when a high flow coefficient is needed. A dome-loaded regulator provides a discharge pressure equal to the pressure signal applied to the dome. Dome pressure

Comp–Day

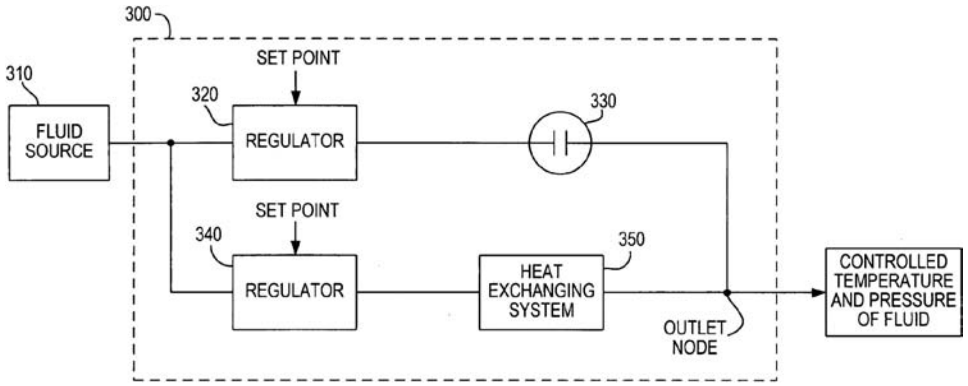


Fig. 3 Method of controlling constant turbine inlet temperature and pressure.
Source: From Ref. 5.

control is provided by a combination of solenoid valves 1050 and 1052 in Fig. 3 that, respectively, either pressurizes the dome with air supplied by an upstream regulator and accumulator, or depressurizes the dome by venting the dome to atmosphere. Operation of these two valves is provided by a digital controller with feedback from downstream sensors. For reliable termination of dome pressure in emergency situations, a redundant NO solenoid valve 1060 is in line with an Emergency Power Off (EPO) circuit and is powered closed during normal operation. Faults such as sensed turbine overspeed or manual EPO will de-power the valve and vent the dome causing the pressure regulator to close (Fig. 4).

THERMAL ENERGY STORAGE

High-temperature thermal storage using 304 stainless steel provides extremely compact, robust, and low-risk TES when designed in accordance with American Society of Mechanical Engineers (ASME) rules. Although material cost is relatively expensive, manufacturing processes are simplified by integrating the thermal storage and pressure retention functions and an annular channel configuration[7] achieves very high heat transfer coefficients and heat extraction efficiencies (Fig. 5).

Advanced micro-porous insulation is used to prevent excess heat loss and provides a temperature gradient of over 600°C with approximately 25-mm thickness. Maximum operating temperature of 700°C is chosen based on life considerations for replaceable cartridge heating elements and creep considerations for the storage material based on estimated time at pressure and temperature. In the case of UPS, the expected cycling is low, but low-cycle fatigue due to dilatation stresses in the inlet piping when cold air from the regulators is introduced were investigated. These were found to be non-issues for UPS. For more extensive cycling that would be required for wind or solar storage, further investigation is needed. In addition, work is ongoing to identify more economical TES designs.

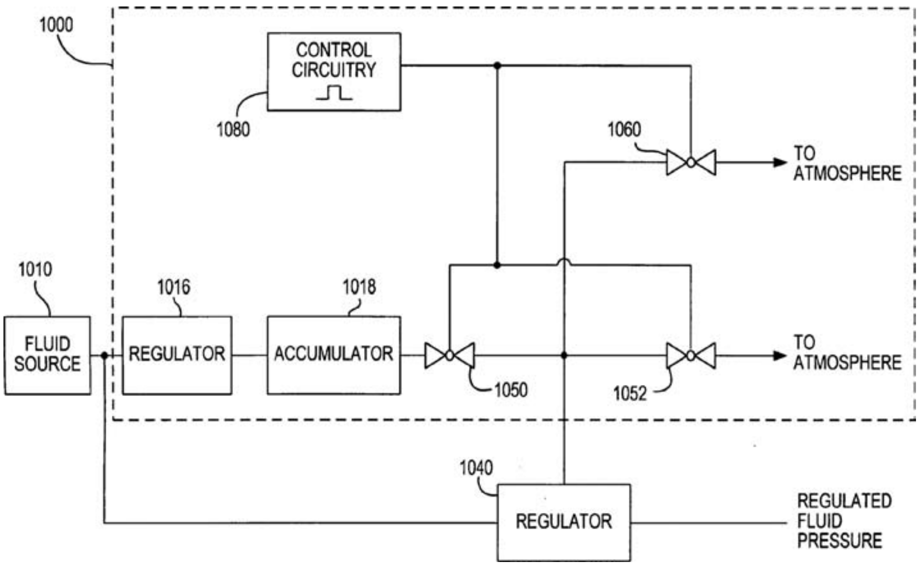


Fig. 4 Method of controlling constant fluid pressure.
Source: From Ref. 6.

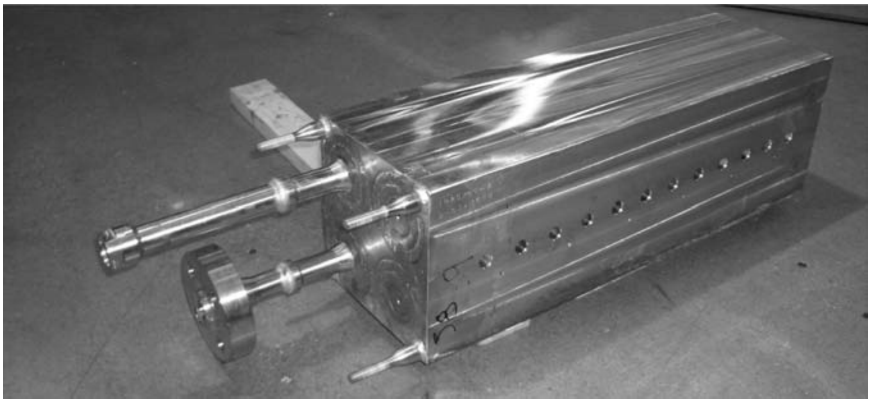


Fig. 5 High energy thermal energy storage for UPS.

COMPRESSED AIR STORAGE IN VESSELS

Most large-scale CAES systems will probably continue to use underground caverns for pressurized air storage where available. However, for small-scale systems or where storage caverns are unavailable, small-storage vessels offer economies of scale due to mass production techniques employed. These cylinders are produced on automated process lines by backward extrusion of billet as opposed to larger vessels produced using seamless pipe with integrally forged heads. One manufacturer claims that the production rate for billet-formed cylinders is less than 30-s per cylinder.

In the United States, transport cylinders having Department of Transportation (DOT) exemption certification for 310-bar [4500-psi] service have been found most economical. Specific costs of less than $0.015 per bar-liter [$0.034 per psi-cu ft] can be obtained, whereas the cost of larger vessels has been found to be twice that or more. For the turbine inlet conditions being employed for UPS, this leads to energy cost for stored air of around $350/kWh. Energy storage costs for compressed air will decrease with increasing turbine inlet temperature. This must be balanced with the cost of the TES needed to achieve higher discharge temperatures.

In the United States, disagreement abounds on regulatory issues surrounding the on-site generation and storage of compressed air. Historically, stationary pressure vessels are designed to ASME standards, and installation and operation are regulated by the individual states. Transport vessels fall under the jurisdiction of the U.S. DOT and are exempted from state control. Many states allow the use of DOT vessels for stationary storage of air, but some do not and some have no regulations on pressure vessels. Attempts to qualify the most economical DOT designs through standard ASME channels have been fruitless since the two organizations' design rules are not harmonized, even though the service seen by the cylinders is comparable. Furthermore, some jurisdictions assess permitting fees on a per vessel basis, whereas others will consider a bank of vessels as one installation and asses a flat fee. Therefore, some users in the United States must pay twice or more for air storage than others.

The regulatory picture for compressed air storage in vessels in Europe is much more favorable with the recent harmonization of the European Union and introduction of the Pressure Equipment Directive (PED). Through appropriate Notified body channels, it is a straightforward matter to re-qualify designs originally intended for one application into another application so long as the types of service are similar.

MODIFICATIONS TO TACAS TECHNOLOGY

One of the most significant adaptations needed to improve TACAS technology for electricity storage is the development of a high-pressure adiabatic compressor and a slight modification to the system architecture to allow heat recovery. A schematic of the proposed architecture is shown in Fig. 6.

The exact form of the high-pressure compressor is under consideration. The high discharge temperatures imposed by adiabatic compression eliminates reciprocating compressors. Large-scale multi-stage centrifugal compressors with direct-drive induction motors have

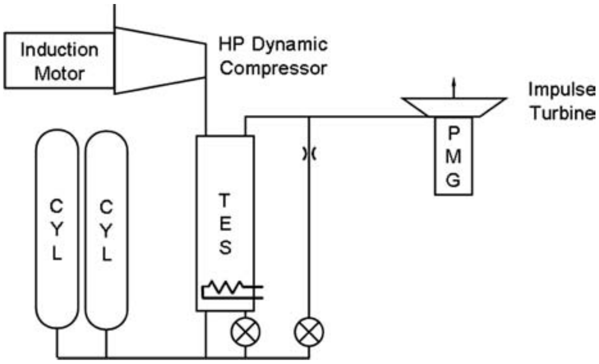


Fig. 6 Schematic arrangement of modifications to thermal and compressed air storage technology architecture for improved cycle efficiency.

been used for high-pressure applications in the oil and gas industry. Scaling dynamic compressor designs for small-scale applications and adapting for high-discharge temperatures poses significant technical challenges, but will be crucial if high cycle efficiency is to be achieved.

A further enhancement to the TACAS technology would allow higher turbine inlet temperatures since limitations on turbine-exhaust temperatures imposed by indoor discharge requirements of UPS could be relaxed. Increasing the discharge temperature reduces the TES discharge temperature range and so must be compensated by higher storage temperatures or additional TES mass for a given output. Further advances in TES material selection and design will be needed to achieve economic viability.

Finally, in applications for solar and wind energy storage, it may be possible to eliminate the bridging flywheel since the turbine has such rapid response. However, the excellent load regulation and speed control attributes of the flywheel in tandem with the turbine make this element of the architecture favorable if power quality enhancements are desired.

SUMMARY

Compressed-air energy storage and TES system have been successfully integrated for UPS applications. Selected features of the TACAS technology are presented and enhancements for improving storage cycle efficiency are discussed. These enhancements include:

- Development of a small-scale high-pressure adiabatic compressor with high discharge temperature capability.
- Modification to thermal storage charge and discharge flow path.
- Turbine bearing upgrades for longer operational life.
- Turbine modifications for increased inlet and discharge temperatures.
- Thermal energy storage cost reduction and or increase in allowable operating temperature.

Focus on these initiatives will allow consideration of TACAS technology for near-term electricity storage in conjunction with wind or solar power generating systems with lower upfront capital expenditure compared to large-scale systems.

REFERENCES

1. Crotogino, F.; Mohmeyer, K.; Scharf, R. *Huntorf CAES: More than 20 Years of Successful Operation*, Proceeding SMRI Spring Meeting, Orlando, FL, U.S.A., 15–18 April 2001.
2. Daly, J.; Loughlin, R.M.; DeCorso, M.; Moen, D.; Davis, L. *CAES—Reduced to Practice*, ASME 2001-GT-397, 2001.
3. Bullogh, C.; Gatzen, C.; Jakiel, C.; Koller, M.; Nowi, A.; Zunft, S. *Advanced Adiabatic Compressed Air Energy Storage for Integration of Wind Energy*, Proceeding of European Wind Energy Conference (EWEC), 22–25 November 2004.
4. Pinkerton, J.F. Thermal and Compressed Air Storage System. U.S. Patent 7086231, 8 August 2006.
5. Schuetze, K.T.; Hudson, R.S. Systems and Methods for Controlling Temperature and Pressure of Fluids. U.S. Patent Application 2006/0076426A1, Filed 5 February 2003.
6. Schuetze, K.T.; Weaver, M.D. Systems and Methods for Controlling Pressure of Fluids. U.S. Patent Application 2006/0060246A1, Filed 17 September 2004.
7. Perkins, D.E.; Hudson, R.S. Thermal Storage Unit and Methods of Using the Same to Heat a Fluid. U.S. Patent 6955050, 18 October 2005.

BIBLIOGRAPHY

1. Vosburgh, K. *Compressed Air Energy Storage*, AIAA Conference on New Options in Energy Technology, San Francisco, CA, 2–4 August , 1977.

Compressed Air Leak Detection and Repair

Robert E. Wilson
ConservAIR Technologies Co., LLP, Kenosha, Wisconsin, U.S.A.

Abstract
Compressed air is a major cost component in manufacturing. As such, it offers one of the largest savings opportunities. The investment in compressing air to energize it and then letting it escape from the system through leaks, without doing any useful work, is a complete waste. This waste can be minimized by implementing a program of leak detection and repairs. This entry covers the subject of how to use a handheld ultrasonic leak detector to locate leaks and the procedures required to implement repairs. The documentation and informational database required to ensure that leak waste is controlled and that new leaks are prevented is summarized. Different application technologies for controlling leaks are presented, and time and cost controls reviewed.

Comp–Day

INTRODUCTION

This entry covers the basic methods of finding and repairing leakages.

The section on Leak Detection presents the most commonly used procedures to identify leakages. The sound signature of air leaks and the use of a handheld ultrasonic acoustic detector to locate leaks are explained. Suggested procedures for tagging and documenting air leaks are presented. The benefit of entering information into a database for historical trending is reviewed.

The section on leak repairs covers the most commonly found air leaks and the procedures to repair them. The logistical problems that create obstacles to expeditious repairs are discussed. The need for after-control, rechecks, and establishing standards are reviewed.

The section on leak control and prevention presents various approaches for managing leaks to minimize waste and ensure that the savings realized are ongoing. The application of flow monitoring and pressure regulation is presented.

Leak costs are discussed and a summary of the entry is presented.

LEAK DETECTION

In a compressed air system, the pressurized air confined by the pipes and vessels escapes from the system through openings as it expands back to atmospheric pressure. Ideally, all of these openings are created intentionally to extract energy from the compressed air in performing a desired task. In reality, however, many of the openings are unintentional, wasteful leaks.

The leak volume is directly proportionate to the area of the opening, the resistance to flow, and the applied pressure differential. The larger the area of the unrestricted opening and the higher the supply air pressure, the greater the leak flow. A chart showing the discharge of air through an orifice is included as an appendix. Note that the values listed are based on a 100% coefficient of flow and should be adjusted for other orifice configurations as suggested (Table 1).

When the air expands back to atmospheric pressure, it transitions from a high-pressure laminar flow to a low-pressure turbulent flow. The escape velocities become extreme as the air volume expands. This results in a full sound spectrum of noise, ranging from audible to high frequency inaudible.

One common method of detecting leakages is to use a soap-like liquid that forms bubbles. Products specifically formulated for high viscosity and film strength exaggerate the bubble effect to enhance the detection capabilities. The liquid is poured, sprayed, squirted, or brushed on a suspect area, and the formation of the bubbles is visually observed. This method allows the detection of leaks that cannot otherwise be heard or felt in the normal operating production environment, but bubble detection is time consuming and messy. It requires the inspection of every connection to the air system, and the foaming agent may require material approval before it can be used in a particular facility. It also is not practical for checking overhead ceiling pipes or under, behind, or inside operating machinery.

The more commonly accepted method for detecting leaks is to use a handheld, ultrasonic acoustic detector that can register the high frequency sound signature associated with gas leaks and translate it into an audible signal. Air leaks have a definitive ultrasonic sound component in their

Keywords: Industrial compressed air system leak detection; Repair; Control and prevention.

Table 1 Discharge of air through an orifice

Area sq. in. press	1/64″ .00019	1/32″ .00077	3/64″ .00173	1/16″ .00307	5/64″ .00479	3/32″ .00690	7/64″ .0094	1/8″ .01227	9/64″ .01553	5/32″ .01973	3/16″ .02761
1	0.028	0.112	0.253	0.450	0.700	1.06	1.48	1.80	2.27	2.80	4.0
2	0.040	0.158	0.356	0.633	0.989	1.42	1.94	2.53	3.20	3.95	5.7
3	0.048	0.194	0.436	0.775	1.25	1.74	2.37	3.10	3.92	4.82	6.9
4	0.056	0.223	0.502	0.892	1.39	2.00	2.73	3.56	4.50	5.55	8.0
5	0.062	0.248	0.560	0.993	1.55	2.23	3.04	3.97	5.02	6.19	8.9
6	0.068	0.272	0.612	1.09	1.70	2.45	3.32	4.34	5.49	6.75	9.8
7	0.073	0.293	0.695	1.17	1.82	2.63	3.58	4.68	5.90	7.29	10.5
8	0.083	0.331	0.741	1.32	2.06	2.96	4.05	5.30	6.70	8.24	11.9
12	0.095	0.379	0.856	1.52	2.37	3.41	4.65	6.07	7.66	9.42	13.6
15	0.105	0.420	0.945	1.68	2.62	3.78	5.15	6.72	8.50	10.48	15.1
20	0.123	0.491	1.100	1.96	3.05	4.40	6.00	7.86	9.92	12.12	17.6
25	0.140	0.562	1.26	2.25	3.50	5.05	6.88	8.98	11.38	13.99	20.2
30	0.158	0.633	1.42	2.53	3.94	5.68	7.7	10.1	12.77	15.70	22.7
35	0.176	0.703	1.58	2.81	4.38	6.31	8.6	11.3	14.26	17.60	25.3
40	0.194	0.774	1.74	3.10	4.84	6.97	9.5	12.4	15.65	19.31	27.9
45	0.211	0.845	1.90	3.38	5.27	7.60	10.3	13.5	17.05	21.00	30.4
50	0.229	0.916	2.06	3.66	5.71	8.22	11.2	14.7	18.60	22.90	32.9
60	0.264	1.06	2.38	4.23	6.60	9.50	12.9	16.9	21.40	26.35	37.9
70	0.300	1.20	2.69	4.79	7.45	10.53	14.7	19.2	24.25	29.90	43.0
80	0.335	1.34	3.01	5.36	8.33	12.04	16.4	21.4	27.10	33.33	48.1
90	0.370	1.48	3.33	5.92	9.25	13.34	18.2	23.7	30.00	36.90	53.0
100	0.406	1.62	3.65	6.49	10.50	14.58	19.9	26.0	32.80	40.50	58.0
110	0.441	1.76	3.96	7.05	11.00	15.82	21.5	28.2	35.60	43.90	63.0
120	0.476	1.91	4.29	7.62	11.40	17.15	23.4	30.5	38.51	47.50	68.0
125	0.494	1.98	4.45	7.90	12.30	17.79	24.2	31.6	40.00	49.25	70.0
150	0.582	2.37	5.31	9.45	14.75	21.20	28.7	37.5	47.45	58.25	84.0
200	0.761	3.10	6.94	12.35	19.15	27.50	37.5	49.0	62.00	76.2	110.0
250	0.935	3.80	8.51	15.18	23.55	34.00	46.2	60.3	76.15	94.0	136.0
300	0.995	4.88	10.95	18.08	28.25	40.55	55.0	71.8	90.6	111.7	161.0
400	1.220	5.98	13.40	23.81	37.10	53.45	72.4	94.5	119.4	147.0	213.0
500	1.519	7.41	16.62	29.55	46.00	66.5	90.0	117.3	148.0	182.5	264.0
750	2.240	10.98	24.60	43.85	66.15	98.5	133.0	174.0	220.0	271.0	392.0
1000	2.985	14.60	32.80	58.21	91.00	130.5	177.0	231.0	291.5	360.0	520.0

Area sq. in. press	7/32″ .03758	1/4″ .04909	9/32″ .06213	5/16″ .07670	3/8″ .11045	7/16″ .15033	1/2″ .19635	9/16″ .24850	5/8″ .30680	3/4″ .44179	7/8″ .60132	1″ .78540
1	5.5	7.2	9.1	11.2	16.2	22.0	28.7	36.3	44.8	64.7	88	115
2	7.7	10.1	12.8	15.8	22.8	31.0	40.5	51.0	63.4	91.2	124	162
3	9.5	12.4	15.7	19.2	27.8	37.8	49.5	62.5	77.0	111.0	152	198
4	10.9	14.3	18.1	22.2	32.1	43.5	57.0	72.0	88.9	128.0	175	228
5	12.2	15.9	20.1	24.7	35.7	48.5	63.5	80.1	99.3	143.0	195	254
6	13.3	17.4	22.0	27.1	39.1	53.0	69.5	87.9	108.0	156.0	213	278
7	14.3	18.7	23.6	29.2	42.2	57.3	75.0	94.7	116.0	168.0	230	300
8	16.2	21.2	26.9	33.0	47.7	64.7	84.7	106.0	132.0	191.0	260	339
12	18.6	24.3	30.7	37.8	54.6	74.1	97.0	122.0	151.0	218.0	297	388
15	20.5	26.9	34.0	41.9	60.5	82.5	108.0	136.0	168.0	242.0	329	430
20	24.0	31.4	39.8	48.8	70.7	96.0	126.0	159.0	196.0	283.0	385	503
25	27.4	35.9	44.5	56.0	80.9	110.0	144.0	182.0	224.0	323.0	440	575
30	31.0	40.5	51.3	63.0	91.1	124.0	162.0	205.0	253.0	365.0	496	618
35	34.5	45.0	57.0	70.0	101.0	137.0	180.0	227.0	281.0	405.0	551	720
40	38.0	49.6	63.0	77.0	112.0	151.0	198.0	250.0	310.0	446.0	607	793
45	41.4	54.1	68.0	84.0	122.0	165.0	216.0	273.0	338.0	487.0	662	865
50	44.9	58.6	74.0	91.0	132.0	180.0	235.0	296.0	365.0	528.0	718	938
60	50.8	67.6	85.0	105.0	152.0	207.0	271.0	342.0	422.0	609.0	828	1,082
70	58.6	76.7	97.0	120.0	173.0	235.0	307.0	388.0	479.0	690.0	939	1,227
80	65.5	85.7	108.0	131.0	193.0	262.0	343.0	433.0	537.0	771.0	1,050	1,371
90	72.3	94.8	120.0	147.0	213.0	289.0	379.0	478.0	592.0	853.0	1,161	1,516
100	79.0	104.0	132.0	162.0	234.0	316.0	415.0	523.0	649.0	934.0	1,272	1,661
110	86.0	113.0	143.0	176.0	254.0	345.0	452.0	570.0	702.0	1,016.0	1,383	1,806
120	93.0	122.0	154.0	190.0	274.0	373.0	488.0	616.0	712.0	1,097.0	1,494	1,951
125	96.0	126.0	160.0	196.0	284.0	386.0	506.0	638.0	789.0	1,138.0	1,549	2,023
150	115.0	150.0	190.0	234.0	338.0	459.0	600.0	758.0	910.0	1,315.0	1,789	2,338
200	150.0	196.0	248.0	305.0	441.0	600.0	784.0	990.0	1,225.0	1,764.0	2,401	3,136
250	184.0	241.0	305.0	376.0	542.0	738.0	964.0	1,218.0	1,508.0	2,169.0	2,952	3,856
300	220.0	287.0	364.0	446.0	646.0	880.0	1,148.0	1,454.0	1,795.0	2,583.0	3,515	4,592
400	289.0	378.0	479.0	590.0	851.0	1,155.0	1,512.0	1,915.0	2,360.0	3,402.0	4,630	6,048
500	358.0	469.0	593.0	730.0	1,055.0	1,430.0	1,876.0	2,360.0	2,930.0	4,221.0	5,745	7,504
750	531.0	696.0	881.0	1,084.0	1,566.0	2,125.0	2,784.0	3,510.0	4,350.0	6,264.0	8,525	11,136
1000	708.0	924.0	1,171.0	1,440.0	2,079.0	2,820.0	3,696.0	4,650.0	5,790.0	8,316.0	11,318	14,784

Table is based on 100% coefficient of flow. For well-rounded orifice, multiply by 0.97. For a sharp-edged orifice, a multiplier of 0.65 will give approximate results. Values calculated by approximate formula proposed by S.A. Moss. $W=0.5303(ACP/\sqrt{T})$; where: W, discharge (lb/s); A, area of orifice (in.^{2}); C, coefficient of flow; P, upstream pressure (PSI, abs.); T, upstream temperature (°F, abs.); Values used in calculating table: $C=1$; $T=530°R(70°)$; $P=$Gage pressure plus 14.7 psi; weights converted to volumes using density factor of 0.07494 lb/ft^3 (correct for dry air at 14.7 psi abs. and 70°F); values from 150 to 1000 psi calculated by Compressed Air Magazine and checked by Test Engineering Dept. of Ingersoll-Rand Co.

noise signature that is beyond the hearing threshold of the human ear. The ultrasonic leak detector translates the ultrasonic noise of the leak signature into an audible sound heard in the earphones worn by the leak surveyor. Some instruments are also equipped with display meters and indicator lights that visually register the magnitude of the air leak. A distinctive, loud rushing sound is produced in the earphones when the leak detector sensor probe is aligned with a leak. With the production background noise suppressed and filtered out by the headphone set, the leakage hissing is heard.

The sound wave generated by an air leak is directional in transmission. The intensity of the leak noise is based upon the shape of the orifice opening, the distance to the sensor probe, and the differential expansion pressure. The sound level is loudest at the actual point of the leakage exit. The procedure for detecting leaks ultrasonically uses this characteristic to locate the actual leaks. Initially, the leak detector is set at the maximum practical sensitivity consistent with the specific environment of the area being inspected. A sweep of the general area is performed as the surveyor walks the system. When a leak is detected, the direction of the leak is determined by scanning the area until the loudest noise level registers. With the probe pointing in the direction of the noise source, the surveyor moves towards the leak, adjusting the sensitivity of the leak detector accordingly. The intensity of the sound increases in the proximity of the leak and is loudest at the actual point of air exit. Extension tubes or cones attached to the sensor probe focus the sound and pinpoint the location of smaller leaks. The bigger, more serious leaks can be felt. A further test using a bubble solution can augment the process by visual enhancement of the exact location. One such product is formulated to produce an ultrasound shockwave as the bubbles burst, so the surveyor gains the benefits of both the visual observation and ultrasonic detection.

Competing sounds often mask a leak or otherwise distort the directional transmission. If possible, the best way to eliminate a competing sound is to shut the system off. If that is not possible, shielding techniques can be applied. The angle of the probe extension can be changed. The competing sound can be blocked using the body or other solid barrier like a piece of cardboard or clipboard. Cupping the hand over the leak, or using a rag, can often isolate the true source of the sound. Bubble tests can pinpoint the location regardless of the competing sound. It is imperative when working in and around operational machinery that safety be most important. Common sense dictates the extent of effort that should be expended to identify and quantify a specific leak.

The first step in the preparation for performing a leak survey is to establish a pattern for surveying the facility to ensure that all the piping, connected use points, and workstations in an area are inspected. Detected leaks are identified and tagged during the surveillance of the system. Different color tags can be used to visually indicate the severity of leaks and establish priorities. Typical classifications might include three levels:

Level 1: Not audible in any environment without an ultrasonic detector.
Level 2: Audible in a quiet environment but not in an operating facility.
Level 3: Serious leaks requiring immediate attention.

Level 1 leaks cannot be felt or heard under any conditions and require the use of the previously described procedures to detect. They are less than 1 scfm and are assigned no value, since the cost of the associated logistics and labor do not economically justify the repair, unless it is very simple, such as the ubiquitous push lock fitting on plastic tubing. Level 1 leaks are tagged and documented for future recheck, since air leaks never fix themselves and only grow larger over time. The cumulative effect of the Level 1 leaks on the compressed air system can be better controlled by maintaining a stable delivered air pressure at the lowest optimum level through the applications of pressure/flow control and regulating use points.

Level 2 leaks are in the 2 scfm range and can typically be felt but not heard without the use of an ultrasonic leak detector. Repairs are economically justifiable and should be performed within a 60-day period.

Level 3 leaks in excess of 2 scfm can typically be felt and sometimes heard by the human ear. These require immediate attention, since they not only waste air but impact the operational efficiency of the compressed air system. Leak flow is a real demand that adds to the filter/dryer loading, increases the pressure drop throughout the system, and creates pressure fluctuations that impact production.

While the true flow for any specific leak cannot be measured practically, the surveyor can assign values based upon the chosen leak volume associated with the various leak levels. These can then be totaled at the end of the survey to estimate the cumulative system leak waste. The surveyor will typically overestimate about the same amount of leakages that are underestimated, so the final figure gives a good portrayal of the total leak waste. As long as the survey procedures are replicated during the re-check, the comparative value for trending becomes an accurate measure for evaluating the remedial repair actions taken. A cost figure can be assigned for use in the financial analysis. Take into account power cost and associated compressor maintenance and repair costs, plus the costs to operate and maintain all the auxiliary equipment, when determining the real value of the leak waste.

Efforts have been made to estimate the actual volume of an air leak based upon pressure and the decibel level registered at a specific distance. People have assembled test stands using the most common orifice configurations

UE SYSTEMS inc.

GUESS-TIMATOR CHART FOR UP9000/10,000

dB vs CFM

DIGITAL READING	100 PSIG	75 PSIG	50 PSIG	25 PSIG	10 PSIG
10 dB	0.5	0.3	0.2	0.1	0.05
20 dB	0.8	0.9	0.5	0.3	0.15
30 dB	1.4	1.1	0.8	0.5	0.4
40 dB	1.7	1.4	1.1	0.8	0.5
50 dB	2.0	2.8	2.2	2.0	1.9
60 dB	3.6	3.0	2.8	2.6	2.3
70 dB	5.2	4.9	3.9	3.4	3.0
80 dB	7.7	6.8	5.6	5.1	3.6
90 dB	8.4	7.7	7.1	6.8	5.3
100 dB	10.6	10.0	9.6	7.3	6.0

NOTES:
ALL READINGS ARE COMPENSATED FOR ATMOSPHERIC PRESSSURE.
All readings were taken at 40 kHz.

PROCEEDURE:

Use the Scanning Module to conduct the broad scanning to pinpoint the air leaks. The Scanning Module with the Rubber Focusing Probe (RFP) is used to determine air losses. The tip of the RFP on the UP9000 should be fifteen (15) inches away from the leak location for determination of the leak rate.

Notice: The values presented in this table are not stated as factual CFM measurement. This table is provided solely for convenience and should only be used as a general guideline.

Factors such as turbulence, leak orifice configuration, pressure, moisture and instrument sensitivity can significantly effect your results.

Fig. 1 Noise vs leak loss at various pressures.

found in compressed air systems, and then have measured air flow and decibel noise at different pressures and distances. One such Chart, published by UE Systems of Elmsford, NY, is presented in Fig. 1. Note the disclaimer that the values are not stated as "factual CFM" and are provided as a "general guideline." A leak signature is affected by many factors, and the loudness of the noise generated is by itself not the sole measure of the volume of the leakage. For example, a high-pitched whistle will sound a lot louder than a low-level whoosh sound, but the whistle will consume less air. At best, the leak detection process will provide an estimate for use in planning the priorities of the remedial repair procedures and a value for evaluating trends.

LEAK REPAIRS

Detected leaks must be visually tagged for future repair. Some tags are configured to enable you to tear off a copy to give to the maintenance supervisor responsible for the leak repairs. Regardless of the configuration of the tag, information sufficient to allow revisiting an individual leak for repair, even if the tag falls off or is missing, should be recorded on a separate worksheet. This typically consists of:

- Recording the unique, sequential tag number assigned to the specific leak.
- Defining its workplace location in a way that is meaningful to the air user.
- Identifying the specific item that is leaking.
- Identifying the actual point of air exit on the leaking item.
- Classifying the degree of leakage so priorities for remedial action are established.

The tag can be used for after-control by providing a place to enter the date and repairperson's name. The supervisor should check to ensure that the repair has been properly completed before signing off and removing the tag. The repair actions and associated time should be recorded on the original worksheet and entered into a database to establish time and cost control accounting procedures.

The surveyor should record complete information on the worksheet to describe the leak. The probable cause of the leak, such as aging, wear, damage, looseness, mishandling, breakage, or other reasons, should be noted with an explanatory note if required. Determine whether the leak should be repaired or a part replaced, and note it on the worksheet. If replacement is recommended, the surveyor should collect enough information about the item to allow for purchasing the repair part or replacement unit. Someone will have to do this if the leak is going to be fixed, so the surveyor should make the extra effort to record the information at the same time the leak is identified. The air user will also need to know if the leak is repairable without having to shut down the associated machinery. The worksheet should have areas for helpful comments and field notes to facilitate remedial actions or to alert people about other issues and opportunities that come to the attention of the surveyor.

Detected leaks must be repaired in order to realize any savings. Since most leaks occur at the operating machinery in the production area, repair procedures tend to be repetitive. Stresses are applied to all the various hoses and couplings, tubing connections, and pipe joints because of machinery vibration and movement of the connected tools and pneumatically driven devices. Over time, leaks develop at sealing areas. These are easily fixed by reconnecting the hose or reinstalling the pipe fitting after inspection and cleaning. Worn couplings or quick disconnects are replaced. The plastic components of point-of-use devises, such as filters, regulators, and lubrications, tend to age and crack over time. These must be replaced. Gaskets and seals dry out and become brittle, so they no longer seal effectively. Valve stem packing and sealing rings, manifold gaskets, hose reel rotary joints, and cylinder shaft seals wear over time and need to be replaced. Clamps, pipe unions, flanges, and pipe groove seals often require re-tightening. Leaks in the compressor room are found around air treatment equipment, condensate drains, receiver manholes, and control tubing.

Leaks on pipe joints are relatively easy to fix by either tightening or reinstalling a connection. Clean all surfaces before reassembly. Use a non-hardening sealing paste for threaded connections to prevent the possible contamination of the air system from torn or frayed Teflon™ tape.[1] Leaks in main headers and branch lines often require lifts or special rigging equipment to gain access, and may require special plumbing skills to repair. Advance planning and scheduling will be necessary for coordinating the repairs on machinery not accessible during production.

The largest obstacle to repairing leaks is the logistics involved in planning and implementing the repair procedures. These logistical problems often take months to resolve and sometimes impede the process entirely. A typical scenario follows.

LOGISTICAL PROCEDURES AND OVERHEAD ASSOCIATED WITH LEAK REPAIRS

1. Meetings and Planning
2. Maintenance requisitions
3. Purchaser—product and supplier identification
4. Order costs—cost per placed order
5. Transportation
6. Control of receipt—administration
7. Storage—space and logistics costs
8. Labor schedule—days/weeks
9. Leakage cost per week/month
10. Time control—verification and administration

LEAK MANAGEMENT

Air leaks grow bigger over time, and repaired leakages usually reappear within six months to one year after they are fixed. Steps must be taken to control the growth rate of leaks and to prevent reoccurrences after repairs are completed. The key to managed leakage control and prevention is rechecks and documentation. Periodic rechecks at predetermined intervals ensure that the leak rate is stabilized at a low level. Through documentation, the trends become obvious and developing patterns, both good and bad, are identified. Problems are recognized before creating issues that are more serious. Taking appropriate actions drives the leak trend downward until it reaches the target established by management, typically 5%–10% of the total air demand. This historical information is used to institute leak prevention measures and for calculating the most economical interval for rechecks to ensure that the gains realized are maintained in the future. Establishing standards and good practices minimizes future leakage. With the time and costs documented, controls can be put in place to properly administer a leak management program.

An alternative approach to implementing a full leak management program is to simply fix the leaks immediately upon discovery, assuming a system is checked for air leaks on a regular basis. The technician brings along a tool tote with the appropriate equipment needed to fix the most commonly found leaks. Usually, only the more serious leaks are addressed in the simple seek and fix approach. Little, if anything, is documented.

The total leakage for a facility can be estimated using techniques that measure pressure degradation over time when there are no production demands on the system.

One such method is to measure the load/unload cycle time of compressors when production is shut down and the only air demand on the system is leakage. Start the compressor(s) and record the on-load time and off-load time over a sampling period long enough to provide a representative average. Calculate the leakage lost as a total percentage of compressor capacity using the formula:

$$\text{Leakage (\%)} = [(T \times 100)/(T + t)]$$

where: T = average on-load time, and t = average off-load time.

In systems configured with compressor controls other than load/unload, leakage can be estimated based upon the total system capacitance. The total estimated volume (V) of all air receivers, the main piping distribution, and other significant air containment vessels must be calculated in cubic feet. Pressure in the main header must be measured at the start and end of the evaluation test period. Production must be shut down so that the only demand on the system is leakage. The compressors are then started in order to pressurize the system to its normal operating pressure (P_1). The compressors are turned off, and the time (T) it takes the system to drop to a pressure equal to half the normal start pressure (P_2) is measured. The leakage is estimated using the formula:

$$\text{Leakage (cfm of free air)} = (V \times (P_1 - P_2)/T \times 14.7) \times 1.25$$

where: V is the volume in cubic feet, P_1 and P_2 are in psig, and T is the time in minutes.

Because air escapes from the system at a rate proportional to the supply pressure, the leak volume rate at the normal start pressure will be much greater than the leak volume rate at the end of the timed cycle when the pressure is half. A 1.25 correction factor is applied to compensate for the difference in the leak rate and to provide a more accurate estimation of the loss.

Installing a flow meter to measure and record the actual flow improves the accuracy of the air leak estimate over using a calculated capacitance based upon estimated system volume. A properly configured flow meter can also be used to monitor the system consumption in order to (1) establish a baseline for evaluating the performance improvements realized from any remedial actions taken, (2) verify trends, and (3) verify that the gains continue to return the investment into the future.

Many of the leaks in an industrial compressed air system are intentional or planned. Condensate drainage and disposal, spot cooling, fume venting and exhausting, material conveying and blowing off, and drying are examples of intentional leaks. Devices are available to eliminate or mitigate the air used to perform these types of assigned tasks.

No air loss condensate drains collect the condensation in a vessel, until it fills with water. A float or sensor detects the high liquid level and opens a drain port, allowing compressed air to displace the water and forcing it to discharge from the vessel. The sensor shuts off the drain port before the vessel is completely drained, so that no air is lost with the water discharge. Some designs are entirely pneumatic, so no electric power is required at the use point. Electrically activated designs require a power source. While this is sometimes inconvenient, electric units have the advantages of (1) indicator lights that show the operational status of the drain and (2) contacts that can be interfaced with a building management system for remote monitoring.

High efficiency blowing devices are available to entrain surrounding ambient air in the primary air stream to increase the impingement force, so that less compressed air is required to perform the equivalent task. Air knives, nozzles, and jets are offered with a variety of different airflow patterns to better suit a specific task. Air amplification ratios as high as 40:1 over open blowing are achievable.[2] Supply pressure at the point of use can often be regulated to a lower level to further reduce compressed air consumption and the associated noise.

Air volume amplifiers are available to create directional air motion in their surroundings and to efficiently move air and light materials. A small amount of compressed air is used as a power source to amplify the flow of entrained ambient air. Airflow is directional, with an inlet and outlet, to exhaust and/or sweep an area in a shaped pattern. Air volume amplifiers can create output flows up to 25 times the compressed air consumption.[2]

LEAK CONTROL AND PREVENTION

Leaks of all sizes, both intentional and unintentional, can be controlled by supplying them, at minimum, an acceptable, delivered air pressure. An air system that has a cumulative equivalent of a 5/16″ leakage orifice, for example, is illustrated in Fig. 2.

The application of Pressure/Flow Control in the compressed air system primary is a good method for minimizing leak waste. The smaller leaks, determined to be uneconomically repairable, leak less at the lower delivered pressure. The Pressure/Flow Control also prevents the

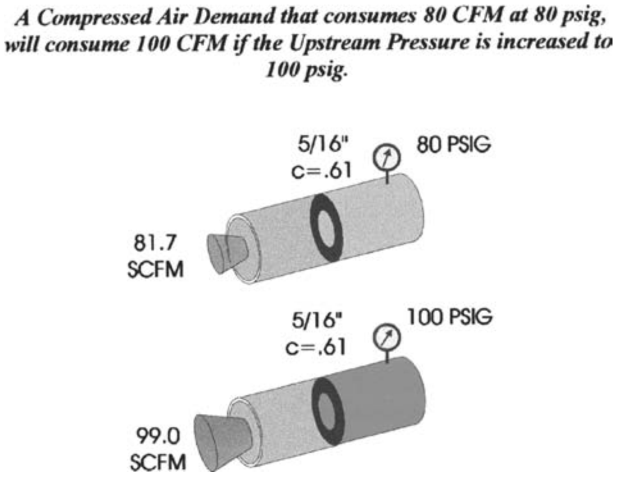


Fig. 2 Cumulative system leak demand at different pressures.

supply pressure from rising because of the lower demand that stems from leak repairs. Without some method of supply side pressure control, the system pressure increases inversely with demand, forcing leaks and other unregulated use points to consume more air. The savings achieved by lowering the leak demand are offset by air that is shunted out elsewhere in the system because of the rising pressure.

The application of point-of-use pressure regulation is another method for minimizing leakages. Setting and securing pressure regulators to supply air at the lowest minimal acceptable pressure maintains the respective leak losses at their lowest possible level.

Shutting off the air to non-productive workstations and assembly lines is another good method of leak control. Lock out valves and isolation valves can be installed to completely shut off the air to machinery that is shut down. The procedure can be manual or automated through the installation of actuated shut off valves that are activated by an external signal. Automation eliminates the dependency on a human action to stop the waste.

LEAK COSTS

The cost of leaks must be determined to allow management to make proper decisions about the compressed air system. The Chart in Fig. 3 illustrates the cost of air consumed by leaks.

In addition to the power cost shown in Fig. 3, consideration should be given to other associated costs, such as labor to log daily operations, scheduled maintenance, repair services, and periodic major overhauls. Leaks are a real demand that require real airflow to satisfy. There is an added cost burden that results from treating the leak air, removing condensation, additional compressor wear, and increased power to compensate for the greater pressure drop because of the higher flow. The final cost figures can more than double the cost based solely upon electrical power.

An effective leak control and prevention program requires continuous monitoring and verification that the gains realized are ongoing into the future. At a minimum, a leak survey should be performed several times a year in a system recheck. Results should be entered into a database and analyzed. Flow monitoring systems are available to measure actual flow. These can interface with management information systems that have remote access. Measuring real flow allows the true cost of the delivered air to be calculated in \$/mmcf (Dollars per million cubic feet). Charting the savings in reports for management ensures continued support for the program.

Air consumed by leaks at 100 psig

Diameter in.	SCFM Leakage	Annual volume		Cost per year*
1/64(.016)	0.41	215,496	cf/yr	\$47.41
1/32(.032)	1.62	851,472	cf/yr	\$187.32
1/16(.063)	6.49	3,411,144	cf/yr	\$750.45
3/32(.094)	14.6	7,673,760	cf/yr	\$1,688.23
1/8(.125)	26.0	13,665,600	cf/yr	\$3,006.43
5/32(.156)	40.5	21,286,800	cf/yr	\$4,683.10
1/4(.250)	113.0	59,392,080	cf/yr	\$13,066.42

**Based upon rate used by US Dept. of Energy, EERE, A Sourcebook for Industry. C = 1.0*

Fig. 3 The typical cost of air leaks.

SUMMARY

In summary, a good leak control and prevention program for a compressed air system starts with a leak survey. Ultrasonic leak detectors are the best tool to find air leaks and pinpoint their location. The information about the air leaks is recorded in a worksheet and documented in a database for use in generating reports and identifying trends. Savings are only realized if leaks are repaired. Repair procedures must be established and an investment made in satisfying all of the logistical obstacles before the actual remedial actions can be taken. Consideration should be given to contracting out the repairs, along with the logistical requirements, to expedite the process and realize the savings as soon as possible. Rechecks and monitoring are necessary to drive the leak trend down and keep it at the targeted rate.

Comp–Day

REFERENCES

1. Ultraprobe™ Instruction Manual Volume 3. UE Systems, Inc., Elmsford, NY.
2. Exair Corporation Catalog No. 20, 11510 Gold Coast Drive, Cincinnati, OH.

BIBLIOGRAPHY

1. ConservAIR Technologies Company, LLP, Kenosha, WI, Industrial Seminar Series, Compressed Air Management.
2. *Improving Compressed Air System Performance a Sourcebook for Industry*, U.S. Department of Energy, Energy Efficiency and Renewable Energy, Prepared for the Compressed Air Challenge® and the United States Department of Energy, Prepared by Lawrence Berkeley National Laboratory, Washington, DC, and Resource Dynamics Corporation, Vienna VA, Section 2–3 *Compressed Air System Leaks*. Nov 2003.
3. Taranto, T. Data Power Inc., *Energy Efficiency in Compressed Air Systems*, 1998, Chapter 6-Compressed Air Systems, Performance.
4. Teflon™ is a trademark of DuPont.
5. Wilson, R. PEMCO Services, St. Petersburg, FL, Presentation, 2005.
6. Winkler, D. LeekSeek International Ltd., Training Presentation, 2005.

Compressed Air Storage and Distribution

Thomas F. Taranto
ConservAIR Technologies, LLP, Baldwinsville, New York, U.S.A.

Abstract
Consistent stable operation of an industrial compressed air system is achieved when compressed air flow supplied to the system equals compressed air demand. Energy distributed to the system is available from two sources; rotating energy of the air compressors, and energy from compressed air storage. Optimum system energy efficiency is possible when the proper amount of energy is available from compressed air storage. Presented here are the physical and mathematical relationships that may be used to assess system performance and determine compressed air storage requirements. These relationships are also applied to design the air storage volume and distribution pressure profile necessary for effective compressed air storage.

INTRODUCTION

Compressed air systems have historically used on-line air compressor capacity sufficient to supply peak air demands. Where air receivers have been installed, the system's pressure profile and lack of storage control limit the effectiveness of compressed air energy storage. Today's energy costs and competitive world economy require that inefficient, wasteful practices of the past must end. This entry develops the necessary calculations to assess compressed air energy demand and calculate the usable compressed air energy in a properly engineered storage system. Methods of design, application, control, and optimization of compressed air energy storage are introduced.

ENERGY FLOW IN COMPRESSED AIR SYSTEMS

Compressed air systems can be viewed as having two parts: supply and demand. The supply system includes air compressors, dryers, filters, and equipment found in the powerhouse/compressor room. The demand side includes perhaps hundreds of use points, including tools, actuators, process use, blowing, cooling, material transport, and sparging (a process whereby air is injected into a tank of liquid, resulting in a bubbling of the solution to provide a desired action). The amount of energy necessary to drive productive air demands is changing constantly as equipment and processes start and stop. During normal machine cycles, compressed air use is often cyclical rather than continuous.

Normal demand variations and diversity of applications result in an airflow profile with peaks and valleys in airflow rates that are at times significantly more or less than the average air demand. For reliable operation of the system, peak air demands must be supplied as they occur. If air supply falls short of demand, system pressure will decrease. When peak air demands are not supplied, system pressure can fall below the minimum acceptable operating pressure. This often leads to lost productivity.

The supply of air is available from two sources; rotating on-line compressed air generation capacity and compressed air energy storage. Operating excessive air compressor capacity to supply peak airflow is inefficient and expensive. Properly engineered compressed air energy storage will supply peak air demand and reduce rotating on-line energy. The result is improved overall system efficiency and reduced power cost.

Compressed Air Generation Efficiency

Generation efficiency is highest when an air compressor is operating at full-load capacity. It is common for compressors to operate at less than full-load capacity, a condition referred to as part-load operation. During part-load operation, a compressor consumes less than full-load power as its compressed air output is reduced. "Specific power" measured as power per unit of air produced (kW/100 scfm), however, is greater during part-load operation. The result is that part-load operation is less efficient than full-load operation. Compressors may also operate in the no-load or unloaded condition. When unloaded, the air compressor produces no compressed air at all yet consumes 25%–30% of full-load power. Running compressors unloaded for a long period greatly reduces a system's overall generation efficiency.

Compressed Air System Efficiency

Supplying peak airflow demand with rotating on-line generation requires one or more compressors to operate in

Keywords: Air; Compressed air; Storage; Receiver; Pneumatic; Supply; Demand; Air storage; Energy; Compressor.

the part-load or unloaded condition. As the peak demand occurs, the compressor(s) will load for a short time during the demand event and then return to part-load or unloaded operation. The result is poor overall system efficiency.

For a compressed air system to achieve maximum operating efficiency, the compressed air supply should incorporate both compressed air generation and storage. The goal is to supply average air demand with generation (on-line rotating energy) and to supply peak airflow requirements from storage (stored compressed air energy).

Compressed Air System Energy Balance

The energy delivered to and consumed in a compressed air system is a function of the weight or mass flow of air moving through the system. The mass of compressed air depends on pressure and temperature. Increasing pressure increases the density and, therefore, the mass of air. Increasing air temperature will decrease the air's density, decreasing the mass of air. This relationship is stated in the Ideal Gas Law.[1]

$$\frac{pV}{T} = \text{a constant} \quad \text{(for a fixed mass of gas)} \tag{1}$$

Eq. 1 Ideal Gas Law for fixed mass of gas.

Compressed air is often measured in terms of its volume—ft^3, for example. The volumetric measure of air is irrelevant with respect to the air mass unless the temperature and pressure of the air volume are also known. Therefore, standards are adopted to express the mass of air under "Standard" conditions, resulting in the definition for a Standard Cubic Foot of air (scf). Standard conditions adopted by CAGI (Compressed Air and Gas Institute) and Compressed Air Challenge® (CAC) are 14.5 psia, 68°F, and 0% relative humidity.

Compressed air energy transfer can be expressed as the mass flow rate of air at a given operating pressure in standard cubic feet per minute (scfm). This is both a measure of volumetric flow rate and the mass or weight flow rate of compressed air. Higher-flow-rate scfm delivers a higher power rate, and the time duration of flow determines the energy transferred.

Compressed air power enters the system from the air compressors and exits the system through air demands, including productive demand, leaks, and all points where compressed air leaves the system, expanding back into the atmosphere. Power delivered from the compressors is measured as mass flow rate Q (scfm) from generation, or Q_{gen}, and power leaving the system also is measured as mass flow rate of air demand, or Q_{dmnd}.

Definition: Q_{gen}. Airflow rate of generation is the compressed air mass flow rate (scfm) produced by the rotating on-line compressor capacity at any moment.

Definition: Q_{dmnd}. Airflow rate of demand is the compressed air mass flow rate (scfm) escaping from the compressed air system to the atmosphere at any moment.

The ideal balance between generation and demand is achieved when $Q_{\text{gen}} = Q_{\text{dmnd}}$.

$$\begin{gathered}\text{Ideal Air System Energy Balance}\\ Q_{\text{gen}} = Q_{\text{dmnd}}\\ \text{Only when system pressure} = \text{constant}\end{gathered} \tag{2}$$

Eq. 2 Ideal compressed air system energy balance.

The first law of thermodynamics for a change in state of a system,[2] or the law of conservation of energy, states that energy is not created or destroyed. This implies that the airflow rate of generation must equal the airflow rate of demand. From practical experience, it is observed that generation and demand are not always equal, resulting in changing system pressure. When $Q_{\text{gen}} > Q_{\text{dmnd}}$, system pressure increases, and when $Q_{\text{gen}} < Q_{\text{dmnd}}$, system pressure decreases. The energy imbalance between generation and demand is either absorbed into or released from storage (Q_{sto}).

$$\begin{gathered}\text{With Compressed Air Entering and Exiting Storage}\\ Q_{\text{gen}} \pm Q_{\text{sto}} = Q_{\text{dmnd}}\\ \text{Entering storage pressure increases}\\ \text{Exiting storage pressure decreases}\end{gathered} \tag{3}$$

Eq. 3 Actual compressed air system energy balance.

Definition: Q_{sys}. Airflow rate of the system is the compressed air mass flow rate (scfm) produced by the rotating on-line compressor capacity (Q_{gen}) at any moment, minus the airflow absorbed into storage ($-Q_{\text{sto}}$ for increasing pressure) or plus airflow released from storage ($+Q_{\text{sto}}$ for decreasing pressure) (Fig. 1).

$$\begin{gathered}\text{Practical Air System Energy Balance}\\ Q_{\text{sys}} = Q_{\text{dmnd}}\\ Q_{\text{sys}} = Q_{\text{gen}} \pm Q_{\text{sto}} = Q_{\text{dmnd}}\\ \text{accounts for changing system pressure}\end{gathered} \tag{4}$$

Eq. 4 Energy balance of compressed air systems.

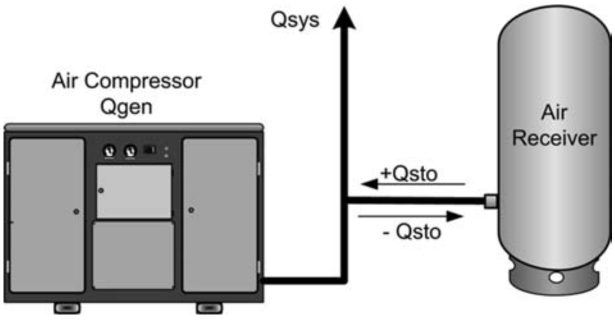


Fig. 1 Airflow relationship: generation, storage, and system flows.

Comp–Day

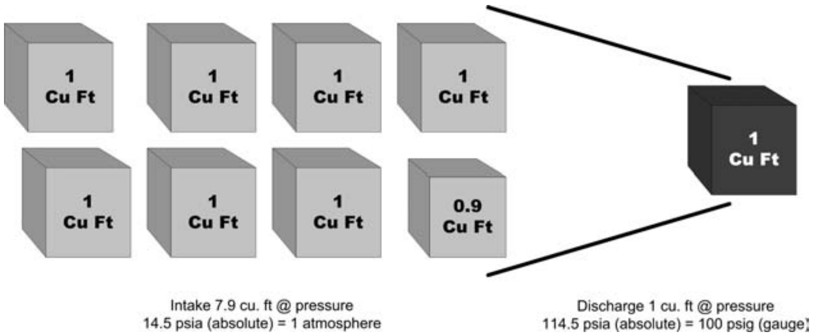


Fig. 2 Compression of air from atmospheric pressure to 100 psi gauge (114.5 psi).

COMPRESSED AIR ENERGY STORAGE

The compressed air system engineer must analyze the dynamics of compressed air energy storage. This section develops the mathematical expressions necessary to study the relationship among air system generation, storage, and demand. To optimize system operation, system energy supply must remain balanced with air demand. Furthermore, system energy supply must be optimized between generation (rotating on-line energy) and storage (stored compressed air energy).

Application of the Combined Gas Law

The combined gas law states that the pressure of an ideal gas multiplied by volume divided by temperature is a constant.

$$P_1\left(\frac{V_1}{T_1}\right) = P_2\left(\frac{V_2}{T_2}\right) \tag{5}$$

Eq. 5 Combined Gas Law.

The compression process usually begins with air at ambient conditions of pressure and temperature (assuming no moisture content). A larger volume of air is compressed to a reduced volume, increasing the pressure of the air. During the compression process, the air's temperature increases. Assuming the air's temperature is ultimately returned to ambient, as a result it can be said that the end pressure of the compression process is equal to the initial pressure times the ratio of beginning volume to ending volume (compression ratio).

$$P_1\frac{V_1}{V_2} = P_2, \qquad \text{where } \frac{V_1}{V_2} = r(\text{Compression Ratio})$$
$$P_1 r = P_2 \tag{6}$$

Eq. 6 Compression ratio.

If during compression, 7.9 ft^3 of air are reduced to a volume of 1 ft^3, the resultant pressure is 7.9 times the initial pressure or the ratio ($r=7.9$ times).

Assume that the compressor intakes 7.9 ft^3 of dry atmospheric air at 14.5 psia. Multiplying by $r=7.9$ results in a pressure of 114.5 psi absolute or 100 psi gauge (Fig. 2).

The application of the combined gas law is the basis of compressed air storage calculation. More advanced forms of the calculation can be used to assess many aspects of compressed air system performance.

If an air receiver of 1 ft^3 volume is pressurized to 114.5 psi absolute, and its discharge valve is opened to the atmosphere, how many ft^3 of air (at atmosphere) will be discharged from the receiver? The form of the ideal gas law used to solve this problem is

$$V_{gas} = V_{rec}\frac{\Delta P_{rec}}{P_{atm}} \tag{7}$$

Eq. 7 Gas volume-receiver volume relationship (assuming temperature = constant).

The volume (at atmosphere) of air (V_{gas}) released from the receiver is equal to the air receiver's volume (V_{rec}) times the pressure change of the receiver (ΔP_{rec}) divided by atmospheric pressure (P_{atm}). Furthermore, the receiver's pressure change (ΔP_{rec}) is equal to the final pressure (P_f) minus initial pressure (P_i) giving ($\Delta P_{rec}=P_f-P_i$). Substituting:

$$V_{gas} = V_{rec}\frac{(P_f - P_i)}{P_{atm}}$$
$$V_{gas} = 1\ \text{ft}^3\,\frac{(14.5\ \text{psia} - 114.5\ \text{psia})}{14.5\ \text{psia}} \tag{8}$$
$$= V_{gas} = -6.89\ \text{ft}^3$$

Eq. 8 Gas volume released from a receiver.

In the previous discussion of air compression, it was shown that reaching a pressure of 114.5 psia requires compressing 7.9 ft^3 of air. The receiver calculations above, however, show that only −6.89 ft^3 of air are released from the 1 ft^3 receiver above.

What happened to the other cubic foot of air? The other cubic foot of air is still inside the air receiver because the receiver's pressure remains at 14.5 psia (1 atm).

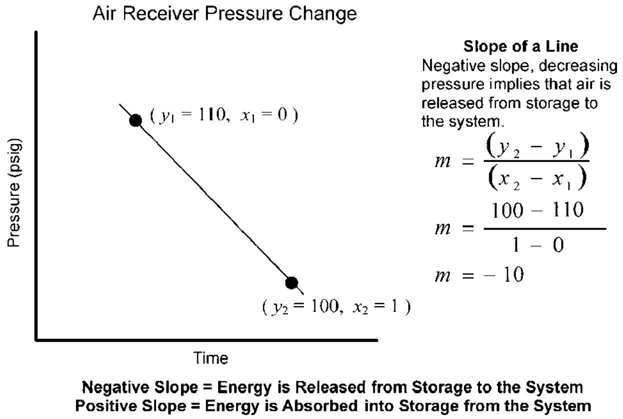


Fig. 3 Air receiver pressure change.

It is important to note that the receiver's pressure change is the slope of the line calculated from final pressure minus initial pressure. Looking at the receiver pressure throughout time, when pressure is falling (negative slope), air is flowing from the receiver to the system. The *xy* plot in Fig. 3 shows pressure (y) and time (x). Storage airflow Q_{sto} has an inverse relationship between the storage receiver and system—that is to say, air leaving the receiver ($-Q_{sto}$) is air entering the system ($+Q_{sto}$).

Pneumatic Capacitance

Pneumatic capacitance of a compressed air system (C_{pn}) represents the compressed air energy absorbed into or released by a compressed air system as its pressure increases or decreases. It is expressed in terms of the mass of air/unit change in pressure—for example, Standard Cubic Foot/Atmosphere (scf/atm) (Table 1).

$$C_{pn} = \frac{V_{sys}}{P_a} = \frac{V_{rec} + V_{pipe}}{P_a} \quad (9)$$

Eq. 9 Pneumatic capacitance.

For every 1 atm change of system pressure, increase or decrease, the system will absorb or release one times its volume (ft^3) of compressed air. Eq. 9 above gives the capacitance of the system in terms of ft^3 per atmosphere of pressure. Assume that the atmosphere is at standard conditions of pressure and temperature, and that the piping volume is negligible.

$$V_{gas} = C_{pn}\Delta P = \frac{V_{sys}}{P_a}\Delta P$$

$$C_{pn} = \frac{100 \text{ cu ft}}{1 \text{ atm}} = 100\,\frac{\text{scf}}{\text{atm}} \quad (10)$$

$$V_{gas} = 100\,\frac{\text{scf}}{\text{atm}} \times 1 \text{ atm} = 100 \text{ scf}$$

Eq. 10 Volume of gas as a function of pneumatic capacitance (C_{pn}).

Eq. 10 shows that a 100 ft^3-volume air receiver changing pressure y 1 atm will displace 100 scf of air into or out of the vessel. The pneumatic capacitance of the system is 100 scf/atm. If the ΔP is positive (i.e., initial pressure is lower than the final pressure), the air displaced is 100 scf, and air is absorbed into the air receiver tank. If the ΔP is negative (i.e., initial pressure is higher than the final pressure), the air displaced (100 scf) is delivered from the air receiver tank.

Because compressed air system pressure is often measured in psi, it is desirable to express the capacitance of compressed air systems in terms of scf per psig (scf/psi).

Table 1 Definition of variables and units of measure

C_{pn}	Pneumatic capacitance (scf/atm) or (scf/psia)
V_{rec}	Receiver volume (cu ft)
V_{pipe}	Piping volume (cu ft)
V_{sys}	System volume (cu ft)
P_a	Atmospheric pressure (psia)
P_i	Initial receiver pressure (psig)
P_f	Final receiver pressure (psig)
ΔP	Storage pressure delta ($P_f - P_i$)
r_s	Storage pressure ratio $(P_f - P_i)/P_a$
V_{gas}	Compressed air volume (scf) standard cubic feet
P_{load}	Compressor load pressure (psig)
P_{unload}	Compressor unload pressure (psig)
Q_{sys}	Airflow rate for the system (scfm)
Q_{gen}	Airflow rate from generation compressor(s) (scfm)
Q_{sto}	Airflow rate of storage (scfm)

Comp–Day

Considering that (P_f psig$-P_i$ psig) yields absolute pressure difference ΔP (psia):Substituting P_f-P_i (psia) for the Storage Pressure Delta (ΔP atm).

$$V_{gas} = \frac{V_{sys}(\text{cu ft})}{P_a(\text{atm})}\Delta P(\text{atm})$$

$$V_{gas} = \frac{V_{sys}(\text{cu ft})}{P_a(\text{psia})}(P_f - P_i)(\text{psia}) \tag{11}$$

$$\text{Therefore}: \quad C_{pn} = \frac{V_{sys}(\text{cu ft})}{P_a(\text{psia})}$$

Eq. 11 Capacitance and storage pressure delta (scf/psia).

Therefore, pneumatic capacitance of a compressed air system (C_{pn}) is a function of the total volume of the system and atmospheric pressure, which can be expressed as:

$$C_{pn} = \frac{V_{sys}}{P_a} = \frac{V_{rec} + V_{pipe}}{P_a} \tag{12}$$

Eq. 12 Capacitance (C_{pn})=scf/psia.

The V_{gas} of the system and the C_{pn} in units of scf/psia are directly related by the change in system pressure (ΔP) in units of psia.

$$V_{gas} = C_{pn}\Delta P \tag{13}$$

Eq. 13 Gas volume as a function of capacitance and delta P.

For the system above with V_{sys}=100 ft^3, the capacitance is 6.896 scf/psia. If the system pressure delta is 14.5 psia (1 atm), the stored V_{gas} is 100 ft^3 (see Eq. 14), because atmosphere (14.5 psia) is the condition defined for Standard Gas Conditions V_{gas}=100 scf.

$$C_{pn} = \frac{V_{sys}}{P_a} = \frac{100 \text{ cu ft}}{14.5 \text{ psia}} = 6.896 \text{ cf/psia}$$

$$V_{gas} = C_{pn}\Delta P$$

$$\text{For } \Delta P = 14.5 \text{ psia(1 atm) the Gas Volume is}:$$

$$V_{gas} = C_{pn}\Delta P = 6.896 \times 14.5$$

$$V_{gas} = 100 \text{ scf} \tag{14}$$

Eq. 14 Calculating gas volume.

Usable Compressed Air Energy in Storage

In the previous discussion, it is apparent that two factors determine the amount of compressed air energy storage: the receiver volume and pressure delta (initial minus final pressure). The volume of a compressed air system is determined primarily by the number and size of air receivers in the system. Piping volume adds to the total, but unless there are several hundred feet of large-diameter pipe, it is often insignificant. For example, 1000 ft of 2-in. schedule 40 pipe has a volume of 23.3 ft^3 or 174 gal (7.48 gal/ft^3), and 1 mi of 1-in. schedule 40 pipe is only 31.7 ft^3 (237 gal). The available pressure delta for storage is determined by the system pressure profile.

Air System Pressure Profile and Storage Delta

The highest pressure available in the system is usually determined by the maximum working pressure of the air compressors. The lowest acceptable operating pressure is determined by manufacturing requirements. With an air compressor rated at 125 psig maximum working pressure and a required use-point pressure of 75 psig, for example, a maximum 50 psig pressure differential is available. Only a portion of this differential can be used for storage, as there are unrecoverable pressure losses as compressed airflows through the system. The pressure profile in Fig. 4 allows for 15-psig control pressure band, 5-psig treatment pressure drop, 2-psig loss through distribution piping, and 8-psig differential in the point-of-use connection piping.

Given a minimum demand-side use-point pressure of 75 psig, and including the imposed pressure delta (10 psi) through distribution plus point-of-use piping, the pressure profile in Fig. 4 shows that the lowest optimum target pressure of the supply-side header is 85 psig. The normal supply-side header pressure is 105 psig. This profile allows for primary storage pressure differential of 20 psig available to the system.

There are costs associated with compressed air energy storage. The discharge pressure at air compressors must be increased to provide storage pressure differential. Increased energy is about 1% for each 2-psig increase in compressor discharge pressure (for positive displacement-type compressors). Also, increasing the compressed air system pressure increases the air demand of the system. Compressed air leaves the system through various openings to the atmosphere, such as the open port of a control valve, a blowing nozzle or open blowing tube, or a leak in the piping. Any opening in the system that does not have a pressure regulator controlling the applied pressure will blow an increased amount of airflow as the system's applied pressure is increased. This increased airflow is called artificial demand. System air demand is increased by approximately 1.0%–1.3% for every 2-psig increase in system pressure. For systems with little effective use-point pressure regulation, artificial demand will be greater.

Definition: Artificial demand. Artificial demand is the additional compressed airflow demand consumed by the system due to actual applied air pressure being greater than the minimum required target pressure.

For the pressure profile shown in Fig. 4, the storage pressure differential is 20–35 psig as compressor controls cycle between their load and unload set points. If the average storage pressure differential is 28 psig, the

Comp–Day

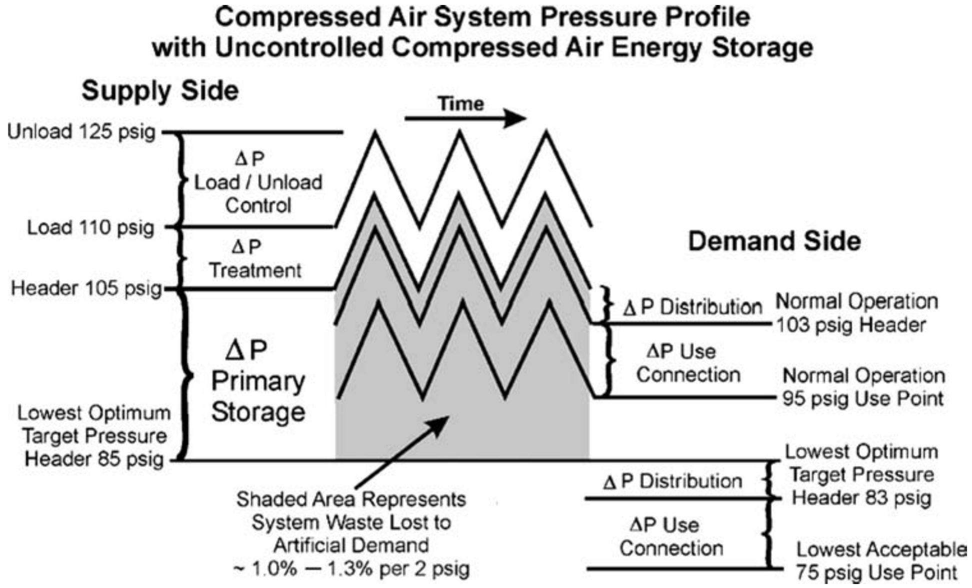


Fig. 4 Compressed air system pressure profile with uncontrolled storage.

resulting power increase is 14%, and waste to artificial demand is between 14 and 18% of the system's airflow.

As shown above, creating compressed air energy storage increases the system's energy requirement and power cost. Therefore, it is unwise to create more storage than the system requires. Proper design will minimize the increased compressor discharge pressure and power requirement. Also, proper control of compressed air energy in storage can virtually eliminate artificial demand. This topic is discussed in "Maximize and Control Compressed Air Energy Storage," presented later in this entry.

But first, the compressed air system engineer must consider various system requirements for stored energy and engineer the storage system appropriately. The following sections demonstrate common applications of the pneumatic capacitance calculation to solve various system energy storage requirements.

Calculate Useable Compressed Air in Storage

Compressed air energy in storage is of use to the air system only if storage pressure is greater than the minimum system supply pressure. Given the pressure profile shown in Fig. 4, and assuming that the air receiver volume of the system is 1000 gal, the amount of usable compressed air energy in storage can be calculated.

For an air receiver of 1000-gal volume with 20 psia primary storage delta P, see Fig. 4; solve for the usable compressed air energy storage V_{gas} (scf) (assuming that the air is at standard temperature and relative humidity).

$$C_{\mathrm{pn}} = \frac{V_{\mathrm{sys}}}{P_{\mathrm{a}}} = \frac{V_{\mathrm{rec}} + V_{\mathrm{pipe}}}{P_{\mathrm{a}}} = \frac{1000(\mathrm{gal})/7.48(\mathrm{gal/cu\ ft})}{14.5(\mathrm{psia})}$$

$$C_{\mathrm{pn}} = \frac{133.67(\mathrm{cu\ ft})}{14.5(\mathrm{psia})} = 9.2(\mathrm{cu\ ft/psia})$$

$$V_{\mathrm{gas}} = C_{\mathrm{pn}}\Delta P = 9.2\left(\frac{\mathrm{cu\ ft}}{\mathrm{psia}}\right)20(\mathrm{psia}) = 184(\mathrm{scf}) \tag{15}$$

Eq. 15 Pneumatic capacitance (C_{pn}) and usable storage (V_{gas}).

Calculating Peak Air Demand

During a large demand event, the supply pressure in the system above is observed to draw down from 105 to 80 psig in 30 s. What is the airflow rate for Q_{sto} (scfm) during the demand event?

Adding time to Eq. 12 for gas volume (scf) allows solving for gas flow rate Q_{sto} (scfm).

$$V_{\mathrm{gas}} = C_{\mathrm{pn}}\Delta P$$

$$Q_{\mathrm{sto}} = C_{\mathrm{pn}}\frac{\mathrm{d}P}{\mathrm{d}T} \quad \text{Where time } T = \text{minutes} \tag{16}$$

Eq. 16 Solve for storage airflow rate Q_{sto} (scfm).

Solving for the peak airflow rate from storage (P_{f}, final pressure; P_{i}, initial pressure),

$$Q_{sto} = C_{pn}\frac{dP}{dT} = C_{pn}\frac{P_f - P_i}{dT}$$

$$Q_{sto} = 9.2\left(\frac{\text{cf ft}}{\text{psia}}\right)\frac{80(\text{psig}) - 105(\text{psig})}{0.5(\text{min})}$$

$$Q_{sto} = -460(\text{scfm})$$

The storage airflow rate is 460 scfm, which is equal to approximately 100 hp of rotating on-line compressor capacity.

Calculating Required Receiver Volume for Demand Events

How much air receiver volume should be added to support the demand event while maintaining supply pressure at 85 psig minimum? Solve Eq. 15 for pneumatic capacitance (C_{pn}); then convert to gal and solve for additional receiver volume.

$$Q_{sto} = C_{pn}\frac{dP}{dT}$$

$$C_{pn} = Q_{sto}\frac{dT}{dP} = Q_{sto}\frac{dT}{P_f - P_i}$$

$$C_{pn} = -460(\text{scfm})\frac{0.5(\text{min})}{(85 - 105)(\text{psia})}$$

$$C_{pn} = 11.5(\text{scf/psia})$$

$$V_{rec} = C_{pn}P_a$$

$$V_{rec} = 11.5\left(\frac{\text{scf}}{\text{psia}}\right)14.5(\text{psia}) = 166.8(\text{cu ft})$$

$$\text{Convert to gallons: } 166.8(\text{cu ft}) \times 7.48\left(\frac{\text{gal}}{\text{cu ft}}\right) = 1248(\text{gal})$$

$$\text{Additional Receiver Volume: } 1248 - 1000 = 248(\text{gal}) \quad (17)$$

Eq. 17 Solve for additional air receiver volume (gal).

Calculating Air Storage for Compressor Permissive Start-up Time

Storage of compressed air energy is also necessary to support compressed air demand during various supply-side events. One common supply-side event is the unanticipated shutdown of an air compressor (due to a motor overload or a high-temperature condition, for example). The startup of reserve compressor capacity requires a period that might range from many seconds to minutes depending on the type of compressors and controls. Most air compressors must start in an unloaded state to allow the electric motor to accelerate to normal running speed. For a typical lubricant-injected rotary screw compressor with part winding or *Y*-Delta, for example, starting might require 5–10 s for transition to full running torque. When the permissive time is past, the compressor's controls must open the inlet to begin compressing air. Then the internal piping, oil sump receiver, and possibly after-cooler must be pressurized before the compressor's internal pressure exceeds the system pressure and the first cubic foot of air is forced through the compressor's discharge check valve into the air system.

Consider the air system shown in Fig. 5, including three compressors operating in a baseload, trim capacity, and standby control configuration. The system air demand is 700 scfm required at 85 psig minimum pressure. The base-load and standby compressors are fixed-speed load/unload compressors with rated capacity of 400 scfm. The trim compressor is a variable-speed drive (VSD) compressor rated at 500 scfm capacity. The VSD trim compressor is set to maintain a target pressure of 90 psig. Assume that the standby compressor is set to start automatically at a pressure of 88 psig and requires a 15 s permissive startup time to deliver its first ft^3 of air into the system.

With the unanticipated shutdown of the base-load compressor, what size air receiver (gallons) is necessary to ensure that the system pressure does not fall below 85 psig during the permissive startup time of the standby compressor?

First, calculate the airflow required from storage after shutdown of the base-load compressor. System air demand is 700 scfm with shutdown of the base-load compressor; the VSD trim compressor will increase its air delivery to full capacity of 500 scfm. The remaining air deficit of 200 scfm must be supplied from storage (Q_{sto}). The pressure profile for the event will result in a fall of pressure to 88 psig before the standby compressor is signaled to start. The minimum pressure for the receiver is 85 psig. Therefore, the initial receiver pressure (P_i) is 88 psig, and the final receiver pressure (P_f) is 88 psig. The permissive startup event duration is 15 s or 0.25 min.

$$C_{pn} = Q_{sto}\frac{dT}{dP} = Q_{sto}\frac{dT}{P_f - P_i}$$

$$C_{pn} = -200(\text{scfm})\frac{0.25(\text{minutes})}{85 - 88(\text{psia})}$$

$$C_{pn} = 16.7(\text{scf/psia})$$

$$V_{rec} = C_{pn}P_a$$

Comp–Day

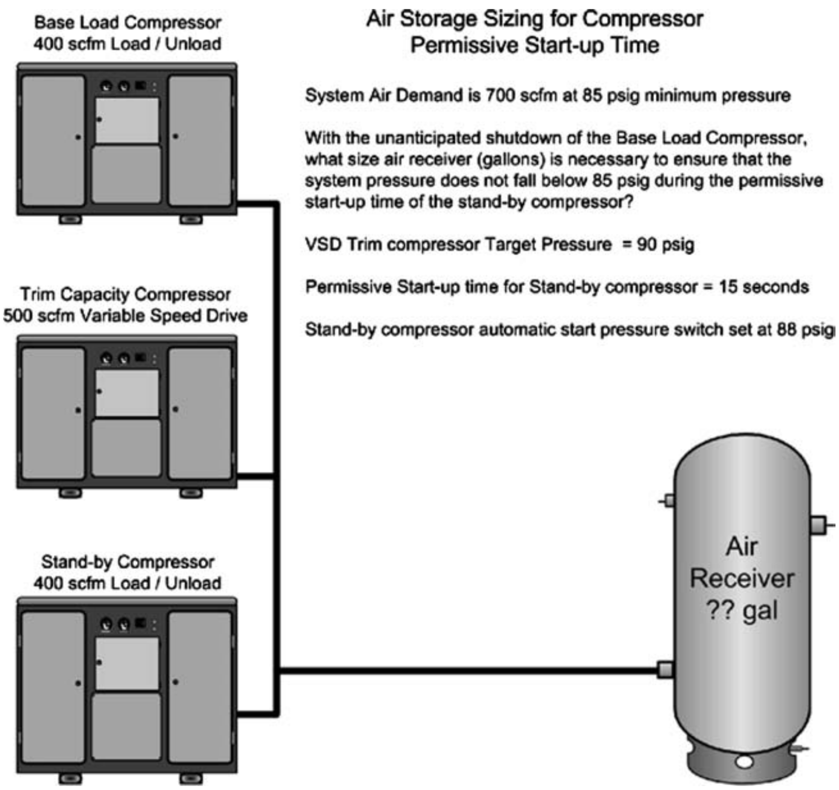


Fig. 5 Base-load, trim capacity, and standby compressed air system.

$$V_{rec} = 16.7\left(\frac{\text{scf}}{\text{psia}}\right)14.5(\text{psia}) = 242.2(\text{cu ft})$$

$$\text{Convert to gallons: } 242.2(\text{cu ft}) \times 7.48\left(\frac{\text{gal}}{\text{cu ft}}\right)$$

$$= 1812(\text{gal}) \qquad (18)$$

Eq. 18 Permissive startup—solve for pneumatic capacitance and receiver volume.

Pneumatic capacitance calculations can be applied to solve a variety of compressed air storage requirements.

MAXIMIZE AND CONTROL COMPRESSED AIR ENERGY STORAGE

For compressed air energy in storage to be effective, the storage pressure must be higher than the demand-side target pressure. As supply-side pressure increases, the power required by positive displacement compressors also increases. The air compressor's power increase is approximately 1% for every 2-psig increase in discharge pressure. Increased storage pressure or increased air receiver volume increases usable air in storage. The economic tradeoff is the capital cost of increased air receiver volume vs the increased compressor supply-side energy cost of higher storage pressure.

Higher pressure also adds energy cost to the system's air demand. Increasing supply-side pressure creates a corresponding demand-side pressure increase. The result is additional energy consumption of the system through an air system loss called artificial demand. Simply stated, if the compressed air pressure applied to leaks and unregulated air use points is increased, the airflow consumed will also increase. Artificial demand is the additional compressed airflow demand consumed by the system when the actual applied air pressure is higher than the minimum required target pressure. Artificial demand in a system without any effective point-of-use pressure regulation will increase the system's energy demand by 2% for each 2-psig increase in pressure. In a "typical" compressed air system, it is common to find that 35%–50% of all air demands have effective pressure regulation. Therefore, artificial demand typically increases by 1.0%–1.3% for every 2-psig increase in applied system pressure.

Artificial demand can be eliminated by controlling the demand-side target pressure at an intermediate point separating the supply and demand sides of the system. An intermediate flow control valve is installed as shown in Fig. 6, downstream of the primary storage air receiver at the beginning of the distribution piping. This separates the supply side from the demand side of the system.

Flow control is used to control the energy (airflow) entering the system while maintaining a real-time energy balance between supply and demand. An intermediate

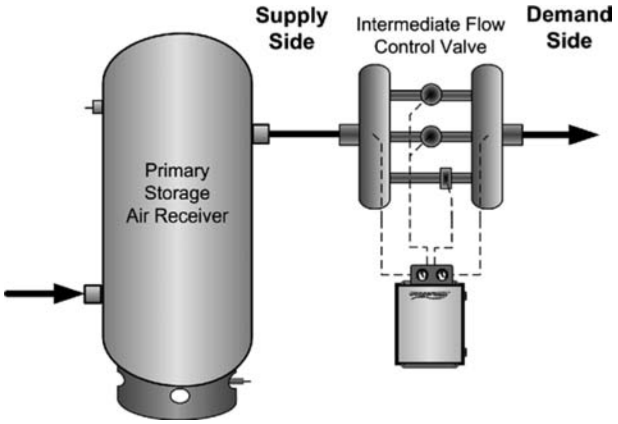


Fig. 6 Intermediate flow control separates supply and demand.

flow control is a packaged assembly of one or more flow control valves in a manifold arrangement with an automatic bypass or fail-safe open override device. It is normally installed in the compressor room at the beginning of the main piping distribution system. Adequate-size receiver(s) installed upstream of the flow control provide compressed air energy storage for controlled release into the system. The flow control senses air pressure at its discharge. Changes in the demand-side energy requirements cause fluctuating pressures. The flow control senses these changes and increases or decreases the airflow from storage as needed to maintain the system's energy balance between supply and demand. The result is stable system pressure set at the lowest optimum target pressure—normally, to within ± 1 psi (Fig. 7).

The upstream compressed air energy storage is crucial to the satisfactory application of an intermediate flow control. The valve package responds immediately to the fluctuating pressure, so compressed air energy from storage must be available for instantaneous expansion into the system to maintain the supply/demand energy balance. The immediate energy supply cannot be dependent on rotating on-line generation. Air must be stored during dwells in the demand cycle when excess compressor capacity is available. Then stored energy is released by the intermediate flow control to satisfy the peak demands.

The intermediate flow control, like all components of a compressed air system, has some unrecoverable pressure loss, which represents an energy cost to the system. Intermediate flow controls are designed to operate with low unrecoverable pressure loss—typically, less than 5 psig. For the pressure profile shown in Fig. 6, the intermediate flow control has an unrecoverable pressure loss of 3 psig, which represents 1.5% increased energy cost at the compressor. In Fig. 4, it is shown that waste to artificial demand is between 14 and 18% of the system's air demand. The net savings achieved by eliminating artificial demand with application of intermediate flow control is 12.5%–16.5% of the system's energy input.

SUMMARY

Compressed air systems constantly undergo dynamic changes in their energy demand. For reliable, consistent, and efficient system operation, energy supply and demand must be balanced in real-time performance. Operating a system without proper energy storage results in part-load or no-load operation of compressors, which decreases system efficiency and increases energy cost.

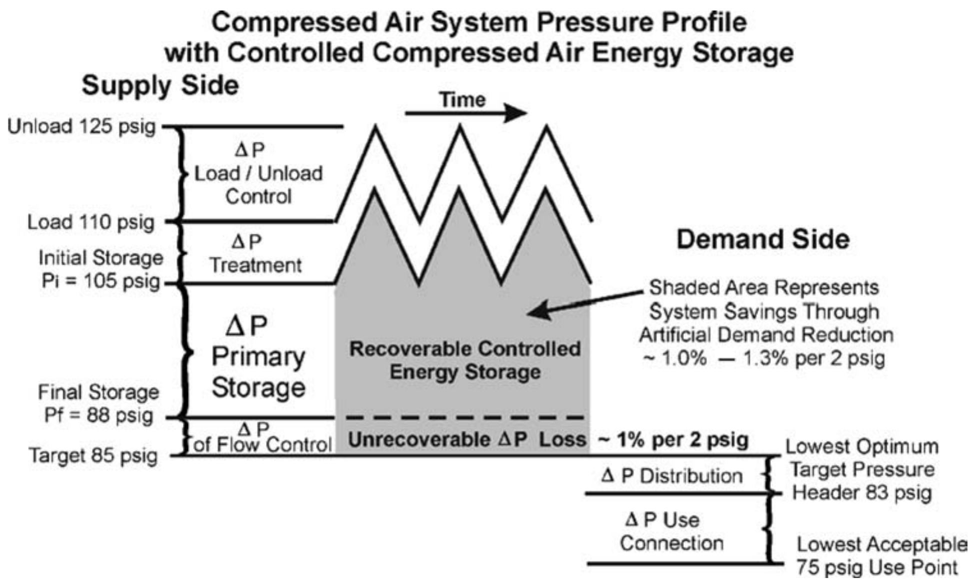


Fig. 7 Compressed air system pressure profile with controlled storage.

Comp–Day

Compressed air energy storage can provide the necessary energy to meet peak demands. Pneumatic capacitance calculations derived from the Ideal Gas Law and First Law of Thermodynamics for systems allow mathematical modeling of compressed air energy storage. Usable air storage is a function of two factors: available storage volume, and the pressure difference between storage pressure and demand-side target pressure.

Demand-side target pressure should be the lowest optimum pressure required to support productive air demands. Increasing storage pressure increases the compressor's energy use by 1% per 2 psig. Uncontrolled compressed air storage pressure also increases the applied demand-side pressure, resulting in waste to artificial demand. Artificial demand typically wastes 1.0%–1.3% of the system's airflow for every 2-psig increase in system pressure.

Artificial demand can be eliminated through the application of intermediate flow control to separate supply and demand. The intermediate flow control paces the energy flow from supply to demand, maintaining a real-time energy balance. The result is reliable, consistent, and stable operation at an appropriate demand-side target pressure.

CONCLUSION

The compressed air system engineer must assess the dynamic energy characteristics of the air system. Compressed air energy storage requirements to support normal system events must be calculated. Energy storage must be optimized with an appropriate system pressure profile allowing the necessary storage pressure differential and adequate air receiver storage volume. Proper control of stored energy and the resultant control response of the system's air compressors, when optimized, will provide the best possible system operating efficiency.

REFERENCES

1. Resnick, R.; Halliday, D. *Kinetic Theory of Gasses: Physics for Students of Science and Engineering*; Wiley: New York, 1966; pp. 57–574.
2. Van Wylen, G.; Sonntag, R. *The First Law of Thermodynamics: Fundamentals of Classical Thermodynamics*, 2nd Ed.; Wiley: New York, 1973; pp. 90–97.

Compressed Air Systems

Diane Schaub
Industrial and Systems Engineering, University of Florida, Gainesville, Florida, U.S.A.

Abstract
Compressed air is a valuable resource for manufacturers, allowing the use of pneumatic-driven hand tools, which can be an ergonomic boon to employees. This resource comes with a price, however, in the form of higher energy costs. This article describes the use of compressed air and the creation and delivery of compressed air from both a supply side and demand side approach. A major focus of this article is on the costs associated with the generation of compressed air and ways to reduce the waste of this resource.

INTRODUCTION

The first section of this entry focuses on the use of compressed air and how it is generated. The section on generation is then separated into a discussion of the supply side and demand side components of a compressed air system. Finally, the costs associated with compressed air, as well as sources of further information, are found at the end of the entry.

OVERVIEW

Compressed air systems could be considered a unique source of energy despite the fact that they are actually powered by electricity.This similarity stems from the fact that compressed air lines can be designed to allow modular tools to plug into the air lines, just like electrical devices can be powered by tapping into electrical outlets.

By far, the most common use of compressed air is to drive pneumatic tools, ranging from nail guns to jackhammers to large drill presses. Pneumatic tools are favored over electric motor driven models because:

- They're smaller, lighter, and more maneuverable.
- They deliver smooth power and are not damaged by overloading.
- They have the ability for infinite, variable speed and torque control, and can reach these very quickly.
- They can be safer because they do not pose the potential hazards of electrical devices, particularly where water and gases are present.

Additional uses for compressed air in the manufacturing sector may include: filtration or control systems, driving conveyors, dehydration, aeration, or refrigeration. As these latter applications do not have the need for portability, and can be performed more cheaply without the additional process step of compressing air, their use is fairly limited. The economics of compressed air will be discussed later in this article.

COMPRESSED AIR SYSTEMS

Although a relatively simple-looking, self-contained air compressor can be purchased at a hardware store, they are limited in size and these small units (battery-powered, gas-powered, or plug-in models) are typically only to be used to fill tires or inflate rafts. Our discussion from this point onward will focus on larger, commercial compressed air systems.

The typical compressed air system is composed of:

- One or more in-series compressors
- An air dryer and air filters
- A receiving tank (for storage)
- Piping
- End uses

Compressed air systems should be perceived as possessing both a supply side and a demand side. Fig. 1 shows a typical block diagram of a industrial compressed air system, with both the supply and demand side noted. These block diagrams are a very helpful first step in understanding how to better manage compressed air systems, as recommended by Ref. 1.

- Improving and maintaining peak compressed air system performance requires addressing both the supply and demand sides, as well as how the two interact in order to have dependable, clean, dry, stable air delivered at the proper pressure. A well-planned balanced system will yield the cheapest and most energy efficient results.

Keywords: Compressor; Pneumatic; Supply side; Demand side; End-use applications; Manufacturing.

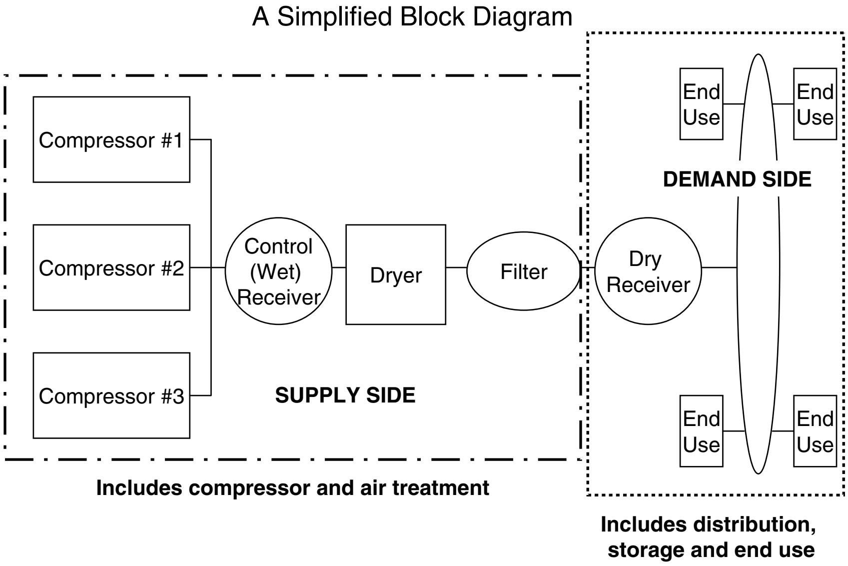


Fig. 1 Schematic of a compressed air system.

COMPONENTS OF A COMPRESSED AIR SYSTEM

Supply Side

A thorough understanding of the end-use compressed air needs, from both a volume and usage profile perspective, is necessary in order to select the appropriate number and size of air compressors. It is rare to find a manufacturing plant that has a constant, uniform use of compressed air throughout the day. Most manufacturing plants have cyclical flow and volume demands due to production schedules, and also desire back-up supply, so engineers typically plan for more than one air compressor to meet a facility's needs. A good strategy is to size a compressor for a base load, and have one or more compressors staged to come online to meet additional compressed air demand. In designing a compressed air system, altitude, inlet air temperature, and relative humidity should be considered, as they impact compressor capacity. More information on how to calculate the influence of these design considerations can be found in Ref. 2, pp. 9–10. It may also be helpful to have different size compressors, so that they can be tailored to fit the operating conditions. Additionally, a small compressor or separate booster may be appropriate for off-shift operations or a special high pressure, periodic application.

The vast majority of industrial compressors are of the rotary screw variety, but double-acting reciprocating or centrifugal compressors are also available for specific applications. Rotary screw compressors come in two configurations: lubricant-injected or lubricant-free. Both have various pros and cons associated with their use. Lubricant-free rotary screw compressors require higher electrical demand, but assure no lubricant carryover. This may be crucial when ultra-clean air is required. On the other hand, lubricant-injected rotary screw compressors have the ability to trim to partial loads to meet usage needs, which can further save on their already lower power costs.

Another issue that can greatly impact the energy efficiency of air compressors is their control strategy. Start/stop, load/unload, and modulating (or throttling) control strategies can be used, depending on the facility's compressed air usage profile.

In order to deliver clean compressed air, filters are installed downstream from the air compressors. The filters remove particulates, some condensate, and lubricant. Regular replacement of filters is necessary to prevent pressure drop, which results in a throttling effect. To illustrate the filter's importance, see the following example:

Example (*Replacement of a Compressed Air Filter Element*)

Assume a 100 hp compressor that operates continuously with an energy cost of seven cents/kWh, resulting in an annual energy cost of $55,328. As the filter becomes clogged, assume the pressure drop increases to six psi across the filter (as compared to a two psi pressure drop for a new filter). Consider that this four psi increase can cost two percent of the annual required energy, or $1100, as compared to $375 for a new filter element.

Another component of a compressed air system is the dryer(s). The compressing of air will condense out the moisture from the natural water vapor found in atmospheric air. This liquid water can cause rust problems in the lines or, should compressed air supply lines connect between buildings, freeze in the winter. Compressed air should be dried to a dew point at least 18°F below the lowest ambient temperature of the demand side.

The various types of dryers are:

- Refrigerated: This is the most common type, with both low initial and operating costs. It can be subject to freezing if operating at low capacities.
- Regenerative desiccant: Typically operated in tandem between two twin dryers, with one operating and the other regenerating. The required volume of purge air needed to regenerate can increase the load or even cause an idle compressor to be started. Heaters can be used in place of purge air, but present their own energy penalty.
- Heat of compression: Similar to the regenerative desiccant dryer, this type of dryer is available for lubricant-free rotary screw compressors and utilizes the hot discharge compressed air to regenerate the desiccant. Their efficiency is affected by changing air temperatures and additional heat may be required for low load situations.
- Deliquescent desiccant: A dissolvable desiccant is used. Regular replacement of this resource is necessary, requiring labor and material costs.
- Membrane-type: A porous membrane separates water vapor from the air and suppresses the dew point. Although there is a low initial cost, these dryers are appropriate only for low-volume applications.

Air receivers can be found on either the supply side (immediately after the compressor or the dryer) or on the demand side, close to the application end use. Air receivers store compressed air and help cover peak events of short duration. If sized properly, they can greatly reduce the frequent loading and unloading of the compressor, saving both energy and maintenance costs. They also stabilize system pressure, which improves performance of the end use.

Other components associated with the supply side may include aftercoolers or intercoolers (for lubricant-free systems), moisture separators, and condensate drains. Depending on the manufacturer, these latter items may be packaged in a single housing with the compressor itself.

Demand Side

Besides a downstream air receiver, the demand side consists of the distribution system or piping, and the end-use applications. Correct sizing of the distribution piping is a critical feature in compressed air system design in order to minimize energy costs.

The piping typically consists of rigid metal or plastic piping from the air compressor room to the general area of the end-use equipment. From this point, flexible rubber or plastic tubing is used, which may be plumbed directly to the end use, or have a shut-off valve with quick-connect attachment points. This flexible tubing may be subject to being run over by foot or equipment traffic and can wear out over time. As a result, air leaks can grow to epidemic proportions, and greatly increase the demand on the compressor. In fact, it isn't unusual to find a poorly maintained system running a compressor that is only feeding leaks. Some facilities will bury large portions of their distribution piping, which make finding and repairing leaks an expensive proposition. A 3/16″ in. hole in a system operating at 100 psig can cost over $5000 a year.

Another operating consideration associated with the demand side is the cost of "normal production." Decisions to add additional applications should undergo a realistic cost evaluation. Consider the following example of an end-use application:

Example (Addition of an End-Use Application)

A quarter inch orifice required to operate a pneumatic hand tools at a recommended pressure of 100 psig was found to have a flow rate of 63.3 scfm (standard cubic feet per minute). After a year of constant use, this equates to 33.3 MMcf (million cubic feet) of compressed air. If compressed air generation costs $300/MMcf, then the power cost for this application will be approximately $10,000/year. If we add additional operating costs of $170/MMcf to account for the operator maintaining the compressed air equipment and the maintenance, lubricant, and repair costs for the system, we find that the cost of this new application use is over $15,000/year. Compare this with less than $2000/year to operate a comparable electrical tool.

High costs can also be incurred through the artificial demand associated with setting the compressor pressure level higher than needed. According to Ref. 3, p. 56, supplying 20% extra psig will force the system to consume 20% more air flow, resulting in 20% waste. Poor applications, such as stuck condensate drains, personnel use of compressed air for cooling or drying, or sparging (aerating of liquids), also use up precious compressed air.

ESTIMATING NECESSARY PRESSURE SET POINT

The determination of the pressure set point for the air compressors needs to be equated. Because of natural pressure drops associated with the components of a compressed air system, as well as unrepaired air leaks, the final point is more difficult to find than just dialing in the pressure recommended by the end-use equipment manufacturer. In fact, it is not unusual for plant personnel to reach the desired pressure by trial and error, increasing the set point until equipment operators stop complaining about low pressure. When possible, pressure measurements should be made after each component of the compressed air system to monitor system performance. Flow or electrical readings can also provide useful performance data. More information on how to calculate optimum compressed air system settings can be found in Ref. 2, p. 205. Fig. 2 provides an example of the pressure drops that can occur along the line.

Estimating Pressure Drop

Measurements can be taken at various points in a compressed air system to monitor the associated pressure drop from each component. The pressure profile shows the lowest pressure seen by the end-uses.

Compressor operating range:	115- 105 psig	Air/Lubricant Separator	5 psid
FRL (Filter, regulator, lubricator)	7 psid	Hose and Disconnects	4 psid
Aftercooler	3 psid	Dryer	4 psid
Filter	3 psid	Distribution System	3 psid

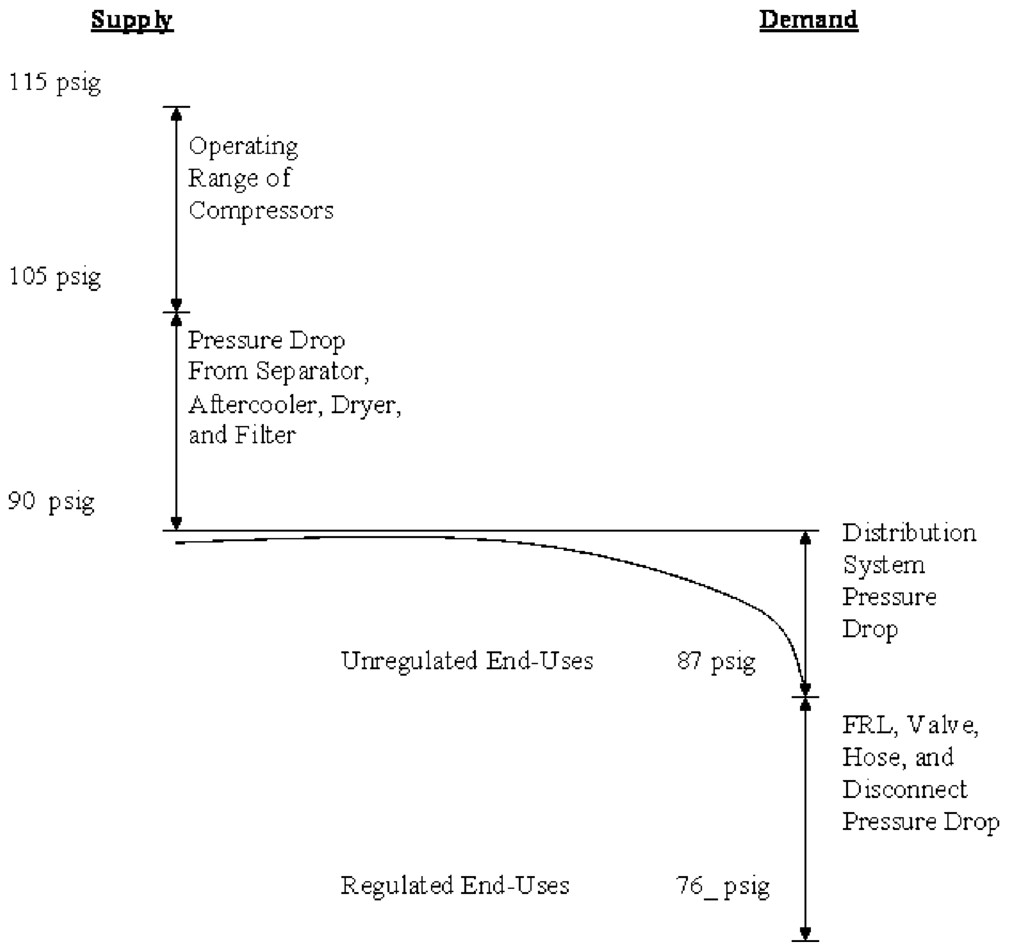


Fig. 2 Estimating compressed air system pressure drop.

COSTS OF COMPRESSED AIR

To operate a one hp air motor, seven to eight hp of electrical energy are required. This large energy penalty, along with the common employee perception that compressed air is essentially a free resource, makes it a challenge to control the costs of compressed air. Inadequate compressor control schemes can cause multiple compressors to run at partial loads, rather than turning them off. Problems with poor maintenance can increase consumption or cause pressure variability. In fact, it isn't unusual to find that compressed air can be the largest end user of electricity.

ARTICLES OF FURTHER INTEREST

The U.S. Department of Energy's Industrial Technologies Program sponsors compressed air training and AirMaster+tools through their Best Practices programs. See http://www.oit.doe.gov/bestpractices for more information. The organization charged with actually delivering the compressed air training can be found at: http://www.compressedairchallenge.org.

REFERENCES

1. *Fundamentals of Compressed Air Systems*; Training Manual for the Compressed Air Challenge, Prepared by Laurel and Associates, Ltd and Resource Dynamics Corporation and presented by the trade association known as the Compressed Air Challenge, (www.compressedairchallenge.org) 1999.
2. Scales, W.; McCulloch, D. *Best Practices for Compressed Air Systems*; The Compressed Air Challenge, 2003; 9–10.
3. Taranto, T. *Compressed Air Management*; ConservAIR Technologies Industrial Seminar Series, Conservair Technologies Company, LLP, Kenosha, WI, (www.conservair.com) 1998; 56.

BIBLIOGRAPHY

1. *Advanced Management of Compressed Air Systems*; Training Manual for the Compressed Air Challenge, 1999; 205.

Comp–Day

Compressed Air Systems: Optimization☆

R. Scot Foss
IR Air Solutions, Davidson, North Carolina, U.S.A.

Abstract
This article will provide you with an complete action plan to optimize your compressed air system including compressor optimization, demand management, density management, and storage in a variety of different applications.

Compressed air represents one of the most critical utilities in most production and process environments. The efficiency of a compressed air system is 100% energy in and, when perfect, produces 11% useful work out. Understanding this, it will cost more to operate a compressor in the first year than it costs to buy and install. Despite this harsh information, power is thrown at symptoms of undefined problems every day. The opportunities of reducing operating cost and energy in air systems is typically more than 50%. This session will carve out a plan of attack to optimize the supply and demand systemically and yield the lowest demand at the highest rate of standard cubic feet per kilowatt of energy.

There are a number of essential actions that need to be taken to optimize the compressed air systems. You need to minimize demand, control the expansion of the air, distribute it while minimizing energy loss, store potential energy, and compress the air efficiently. Other than operating the compressors, as efficiently as possible, everything else seems to elude most everyone. This work cannot be done theoretically on each piece of equipment only in the compressor room. It must be done systemically. More efficient compressors make more air with the same amount of power. They cost more and can be an important part of a well-operated system. On the other hand, if you throw a more efficient compressor at a highly inefficient system, you will waste more air at the same operating cost and save nothing.

CONTROLLING DEMAND IN THE SYSTEM

1. Control the expansion of the compressed air to the point of use. You must control 100% of users with regulation, which is adjusted lower than the lowest supply pressure. If it is not possible to achieve this with operator discipline, then you must use a demand controller or expander at a central location adjusted in the same manner.
2. Reduce the pressure differentials on installation components such as filters, regulators, lubricators, tube, hose, and disconnects on the demand side of the system. The intent is to operate demand at the lowest possible supply pressure on critical high-pressure applications.
3. Flat line the high rate of flow, intermittent applications with dedicated storage and metered recovery. This is much like a battery charger or water tower application. This can also be a pressure driver for the operating protocol. You will slightly increase the base usage and eliminate peaks.
4. Review and add as necessary general and control storage to slow the rate of change in the system. This will allow you to maintain a higher point of use pressure if necessary without increasing the supply pressure. If there is any diligence used, you can normally reduce the supply pressure simultaneously.
5. Upgrade the quality of information to track progress and improve decision making. This should include a flow meter and demand pressure monitor at the discharge of the demand controller or the expander. If you do not use a demand controller, recognize that demand is only accurately displayed when the demand exceeds the supply. This is referred to as a negative rate of change. When supply exceeds demand, which is a positive rate of change, you are measuring supply response to demand. The system will take whatever supply power you throw at it. A 450 scfm negative rate of change will recover to the original pressure in 1 min, if we respond with a 200 hp compressor. If we throw a 400 hp compressor at the event, it will recover in 15 s at a more rapid rise in pressure. The inefficiency is the part load energy of the larger compressor for the balance of the 45 s. If we match

☆ This entry originally appeared as "Optimizing the Compressed Air System" in *Energy Engineering*, Vol. 102, No. 4, 2005. Reprinted with permission from AEE/Fairmont Press.

Keywords: Compressors; Compressed air; Storage; Demand controls; Expanders; Metered storage; Leaks; Transient events; Air dryers and filters; Potential energy.

Comp–Day

the event with a 100 hp compressor, the pressure will hold at the load pressure of the compressor until the event stops, at which time, the pressure will recover at the same rate of rise as the initial rate of decay.

6. Review and add as necessary general and control storage to slow the rate of change in the system. This will allow you to maintain a higher point of use pressure, if necessary, without increasing the supply pressure. If there is any diligence used, you can normally reduce the supply pressure simultaneously.

REDUCING DEMAND IN THE SYSTEM

1. Develop a leak benchmarking program on a gradual reduction of the tolerance volume. Select a level at a known low load, and repair your way to that level. Every several weeks, check the low load and scan the system using an ultra sonic leak detector. Find and repair the largest leaks found to bring the system back into benchmark. When you are comfortable with this level, lower the level and begin again. You will reach a point where there are so many small leaks to fix during the benchmarking period, the labor hours cannot be justified. At this point return to the previous higher tolerance value.

   Record the types and nature of the leaks that you are fixing, so that you can leverage this information into buying more leak resistant components and improving best practices installations. Note that it is important that the reduction of demand does not cause the demand pressure to rise. If it does, then other unregulated users will increase at the elevated pressure. That is why it is so important to have demand controls installed before you become aggressive in demand reduction. It is also important to off load a linear amount of supply energy for the demand reductions.
2. Eliminate all open compressed air blowing applications and replace with low pressure centrifugal or positive displacement blowers, if at all possible. If it is not possible to use blowers, apply specialty air volume reducing nozzles for the application. Take your time with these applications developing the thrust per square inch as close as possible to the open blowing application. You will also need to filter the air for specialty nozzles, as they will easily plug up with pipe debris. Whenever possible, use a solenoid valve to shut of the air on cyclical applications.
3. Replace all applications, which are poor users of compressed air. Focus on operating cost alternatives. Use electricity whenever possible for its better wire to work energy relationship.
4. Reduce the size of demand events as seen by the system including high ramp applications. This can be accomplished by slowing down the introduction of these events into the system. This can be done by opening the demand valve slower manually or automatically. This reduces the "ramp in" rate of flow, so that the supply including control storage can match the event limiting the ultimate pressure drop, which would result.
5. Regulate all points of use, even if you have installed a demand controller or expander in the main supply system's piping. Make sure that the set points on the regulators are equal to the minimum supply pressure minus the point of use filter and regulator pressure drop or less. If you allow for a 2–3 psig margin below this value, small leaks and filter dirt loading will not cause frequent changes in process performance.
6. Limit the coincidence of events that cause peak demands in the system. This includes minimizing the blow duration on timer drains and adjusting intervals seasonally for relative humidity. Move large events to low load times where possible.
7. Shut off all air using equipment when not in use. Make sure that the shut off valves are ergonomically installed, so that operators can easily reach them. If this does not work, install solenoid shut off valves that are tied into the electrical shut off on the machine, work station, or process.

STORE POTENTIAL ENERGY TO SUPPORT TRANSIENT EVENTS INCLUDING A COMPRESSOR FAILURE IN THE SUPPLY SYSTEM

1. Convert enough kinetic energy to potential energy so that you can handle largest event without turning on another compressor during normal operation. If you do this, you will also handle all of the smaller transient events that are not controlled from downstream. This can include the coincidental impact of a third to first shift startup. Remember that storage is a function of the capacity to store air times the useful differential across it. If you are operating constant pressure compressor controls and they operate correctly, no amount of capacitance will generate any useful storage.
2. Store enough air on the supply side of the system to manage a desired pressure drop, while bringing up a backup compressor to replace a failed one. The intent would be that the event will have no impact

Comp–Day

on the process or production serviced by the system. The intent is to operate only the supply that is required at any time with everything else off.

Example: largest compressor = 1600 scfm, maximum allowable pressure drop from the load pressure on the back up compressor = 10 psid, permissive time to load the compressor from a cold start signal to full load = 15 s, atmospheric pressure = 14.3 psia, gallons per standard cubic feet = 7.48 gal

$$1600 \times (15/60) \times (14.3 \times /10) \times 7.48$$

$$= 4278.6 \text{ gal}$$

3. Create enough storage to control the maximum load cycles per time period on any trim compressor. It is safe to say that 3 min load–unload cycles or longer would be desirable on any positive displacement compressor. This can get trickier on large dynamic compressors, but it is not impossible.
4. If the size of any event or compressor is too large to handle with control storage or you want to protect the system and production against an electrical outage, single phase, or brown out, offline high-pressure peak shaving would be the most desirable approach to minimize on board power. It would not be unusual to store 30–40,000 ft^3 of air in a 100 psig differential supported by a 20 hp compressor offline. You would then introduce the air back into the system on variety of different cues or logic patterns to support the various events.

   Note that it is the intent of all potential energy applications to either prevent the normal operation of an additional compressor, extend the mechanical life of a compressor or compressors, or both. Well applied storage will increase the base load in the system slightly, and eliminate the requirement for added compressors during peak plus the inefficient part load in between peaks.

DISTRIBUTE THE COMPRESSED AIR, WHILE MINIMIZING ENERGY LOSSES

1. The concept of design or redesign should be to minimize the highest amount of air mass or volume of air and the distance that the air must flow to support any part of the system from supply to demand.
2. Resistance to flow is necessary in the system. Without resistance to flow there is no flow. As the system is open on both ends of the system all of the time to a larger or lesser degree, resistance to flow and storage keeps it functioning. Mass flow restrictions are differential pressures in the system, which change as a square function of flow change. It is important to design or retrofit your system for a maximum differential at highest flow, highest temperature, and lowest inlet pressure. This will produce the highest differential pressure across the components being evaluated. Although we are recommending a conservative approach towards this process, the piping distribution system should not be made intentionally oversized or all the same size for convenience. Oversized piping will not provide economical storage and will make it difficult for supply to see demand efficiently. A reasonable differential pressure would be 1–2 psid from the discharge of the cleanup equipment at the supply or the discharge of the demand controller, as it applies to your system, to the farthest point in the demand system at the previously discussed design conditions.
3. In most systems that have distribution problems, you should minimize waste and flat line transient users with dedicated storage and metered recovery at the point of use before considering making changes in the piping distribution system.

As little as a 10%–20% demand reduction at the peak condition can be sufficient to eliminate the most distribution losses and the requirement for piping retrofits.

REDUCE SUPPLY ENERGY WHEREVER POSSIBLE

1. When 100% of demand is at a lower pressure than the lowest supply pressure, set up the supply pressure to optimize the pound per kilowatt of compressed air energy for the on board compressors. Operate all compressors that need to be on flat out and optimized except one compressor trimming and all other compressors off regardless of inlet conditions or relative demand load. You must optimize the compressor and the motor simultaneously. Optimal means the most pounds or standard cubic feet at the optimal density (pressure and temperature), while managing the highest power factor and motor efficiency simultaneously. In this scenario, the trim compressor is the only compromise to "optimal" assuming you can maintain a range of supply pressure across the range of load conditions that relates to optimal on the base load compressors. Another option is to trim with variable frequency drive compressors using storage continuously, while adding and subtracting base load compressors. The Variable

frequency drive (VFD) compressor or compressors will displace or fill in the removal or addition of a base. In this case, you will optimize both the base load compressors and the trim compressors at the same time.

Note that this is called a "Bellows Effect" operating protocol.

2. Base compressors should always be selected based on the best energy efficiency. Trim compressors should be selected first on operating speed to cold or hot start and shut off capabilities, and secondly on their flexibility for automation interface. If you are trimming with VFD/s, the same requirements are applicable. This typically translates into smaller, less permissive compressors. You must be certain that the total trim capacity (one, two, or three trim compressors) is equal to or larger in capacity than the largest base compressor in the supply arrangement. This will assure that there are no gaping holes in supply, so that you can make smooth transitions from one power level to the next. Supply systems that do not have this capability end up running too much power part loaded all of the time to support the transitions. Remember that bigger is more expensive.
3. Develop an operating profile for the supply system, which optimizes the compressors based on a full range of usage and conditions. In most systems, the only time the system is remotely efficient is during peak load. It generally goes down hill during lighter or low load. Also evaluate the full range of system's usage against the full range of ambient inlet and cooling conditions to determine how the system will work before you make any final plans on equipment selection. Make every attempt to manage peaks with potential energy instead of on line power. You must also evaluate the risk of a unit failure in order to have a solid curtailment plan. If brown outs or black outs are common, you must include this in your plan.
4. Unload all unnecessary ancillary power, such as dryers, pumps, fans, etc. through the use of more efficient controls and motor drivers. Size all filtration and dryer equipment for a total differential of 3.5–5 psid. The differential should be at the highest inlet flow, highest inlet temperature, and the lowest inlet pressure. The differential on the filtration should be in a wet and clean condition. Plan the additional differential from dirt loading when selecting the compression equipment, so that you do not overload the motor drives as you will absorb the added differential at the air end discharge. We would recommend no more than an additional 1.5–2 psid on the total filters. There are filters available to accomplish this with a change every 5–6 years at this dirt loading rate. The total differential across all cleanup equipment should not influence the total connected horsepower on the compressors by more than 4% at the worst case maintenance condition.
5. Use a master signal for the compressors located in the dry clean storage downstream of the contaminant control equipment. If the signals are in the compressors upstream of the cleanup equipment, the compressors controls will respond to the demand interpreted through the differential pressure, which changes as a square function of flow change. This causes the compressors to over shot and under shot, which results in hunting. This requires excess energy to compensate.

   Please note in the illustration that we have installed a three-way valve so that you can return to local control signals when you wish to isolate the compressor from the system. It is also important to note that the adjustment of the compressor controls, with a master control signal, should be based on controlling downstream of the cleanup. If the pressure across the cleanup equipment is 10 psid, when you moved the signal, you would also want to reduce the control set points on the compressor/s an additional 10 psid. This is because you will absorb the differential when you move the signal and without adjusting the operating set points for the compressor/s, you may overload the motor.
6. Develop an operating profile which takes control storage, set points of the compressors, signal locations, and differentials into account. Put it down on paper prior to implementing it and check the range of conditions to make sure it will work. Do not put fudge factors into the profile. This is not an art form. It is a science. If you are not sure of what you are doing, contact a technology firm who can assist you. Literally, 95% of all compressor profiles are not set up correctly. Most engineering firms that design systems select the equipment and never think through the operating protocol or profile prior to installation.
7. Finally, you must get the system to operate effectively and efficiently before you automate it. More than 90% of the time, users try to apply automation to a system to get it to work properly. If you automate a system that does not work, you will have an automated mess. You must be able to get it to work correctly on the local controls first. When and if you automate, keep in mind that their purpose is to refine the operating cost and reliability issues across all conditions unattended. Automate the operation based on at least rate of change, storage, time, and pressure. You may even wish to add a selective rate of change protocol, which chooses the correct compressor for the situation. Take your time and test your concept

prior to making the decision by preparing algorithms including transitions of power and demand including failure scenarios. Keep in mind that you do not have to match the event in the system. You only need to slow it down so you can wait longer. The essence of a masterfully designed system is the ability to control demand by matching transient events as quickly as possible with an expander or demand controller serviced with potential energy.

Once this is accomplished, the compressors' control job is managing control storage by replenishing it as slowly as possible. The longer you can take, the less energy you will use.

SUMMARY

A compressed air system is a highly interactive configuration with all aspects affecting all other aspects. Developing an action plan to improve the efficiency and reduce the operating cost can be rewarding, but must be done in the correct order to enjoy the success and avoid production inconvenience. It is a process of black and white with a lot of gray in between. Far too many owners want to buy a solution, rather than apply one. Problem definition, metrology, and carefully planning are all essential. When you have completed the action plan, do not forget to measure the results. Validation is necessary to support you return on investment strategy.

Cooling Towers

Ruth Mossad
Faculty of Engineering and Surveying, University of Southern Queensland, Toowoomba, Queensland, Australia

Abstract
Cooling is necessary to many industrial processes, such as power generation units; refrigeration and air conditioning plants; and the manufacturing, chemical, petrochemical, and petroleum industries. As recently as 20 years ago, cooling towers were more the exception than the rule in the industry because of their high operating cost and the large capital required for their construction. Due to the recent stringent environmental protections, cooling towers became more common. Cooling towers range in sizes and types. Wet, dry, and hybrid are the main types, and each type has many variations in design according to the way the fluids are moved through the system. Some of the advantages and disadvantages of these types, methods of determining their performance, and some terminology common to the cooling industry are presented in this entry.

INTRODUCTION

Most industrial production processes need cooling of the working fluid to operate efficiently and safely. Refineries, steel mills, petrochemical manufacturing plants, electric utilities, and paper mills all rely heavily on equipment or processes that require efficient temperature control. Cooling water systems control these temperatures by transferring heat from hot process fluids into cooling water. Through this process, the cooling water itself gets hot, and before it can be used again, it must either be cooled or be replaced by a fresh supply of cool water.

A cooling system in which the water used in cooling processes or equipment is discharged to waste is called once-through cooling. Characteristically, it involves large volumes of water and small increases in water temperature. Once-through cooling is usually employed when water is readily available in large volume at low cost. Common sources are rivers, lakes, and wells, where the only cost involved is that of pumping. But with today's need for water conservation and minimal environmental impact, industry is turning more and more to recycling water in what are called cooling towers.

Recently, cooling towers are becoming widely used in most industrial power generation units; refrigeration and air conditioning plants; and the manufacturing, chemical, petrochemical, and petroleum industries to discard waste heat to the environment. They range in sizes—the smallest cooling towers are designed to handle water streams of only a few litres of water per minute supplied in small pipes like those in a residence, whereas the largest cool hundreds of thousands of litres per minute supplied in pipes as huge as 5 m in diameter in a large power plant.

Cooling towers are believed to be only the direct contact type heat exchangers. They can be direct or indirect, however, and they are also characterized in many other different ways based on the type of fluid being used in the cooling process, the means by which the fluids are moved, and the way the two fluids (hot and cold) move with respect to each other. Description of these types, some of their advantages and disadvantages, and methods of estimation of their performance are discussed in the following sections. Some useful terms common to cooling towers industry are also given. This information is a collection of materials published in the list of references and Web sites at the end of this article.

TYPES OF COOLING TOWERS

Cooling towers are classified mainly on the basis of the type of fluid used in the cooling process—water or air. These are three types: wet (evaporative), dry (non-evaporative), and wet–dry (called hybrid).

Wet Cooling Towers

The wet type is more common in large cooling towers, such as in electrical power generation. It is a direct contact heat exchanger, in which hot water from the condenser and cooling air come into direct contact. The water flows in either open circuit or closed circuit. In open circuit, cooling water is pumped into a system of pipes, nozzles, and sprayers within the tower, and is drawn by gravity into a pond below (Fig. 1). Air from the atmosphere enters the tower from the bottom of the tower and flows upward

Keywords: Cooling; Towers; Wet cooling; Dry cooling; Hybrid cooling; Plume; Performance; Counter-flow; Cross-flow.

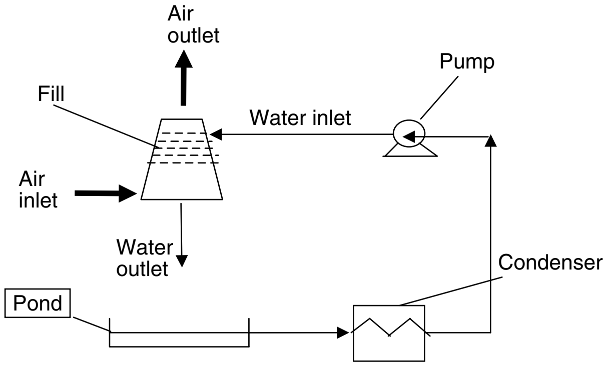


Fig. 1 Open-circuit cooling tower.

through the falling water. The two fluids go through a material that is provided to increase the surface area of contact between them, which is called packing (or fill). The heated and moisture-laden air leaving the fill is discharged to the atmosphere at a point remote enough from the air inlets to prevent it from being drawn back into the cooling tower. The water is collected at the bottom of the tower and then recirculated to remove more heat from the condenser. The temperature of the cold water entering the condenser will determine the steam condensate temperature and, hence, the backpressure, which impacts the efficiency of the whole power generation system.

The closed-circuit cooling tower (Fig. 2) involves no direct contact of the air and the liquid—usually, water or a glycol mixture—that is being cooled. This cooling tower has two separate fluid circuits. Water is recirculated in an external circuit outside a closed circuit made of tube bundles or coils containing the hot fluid being cooled. Air is drawn through the recirculating water cascading over the outside of the hot tubes, providing evaporative cooling similar to an open cooling tower. In operation, the heat flows from the internal fluid circuit, through the tube walls of the coils, to the external circuit and then (by heating of

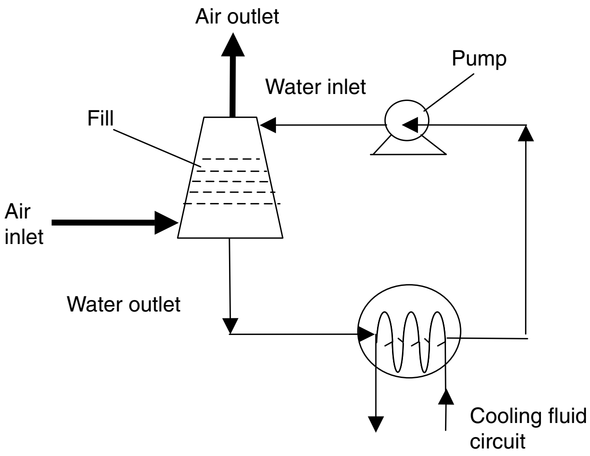


Fig. 2 Closed-circuit cooling tower.

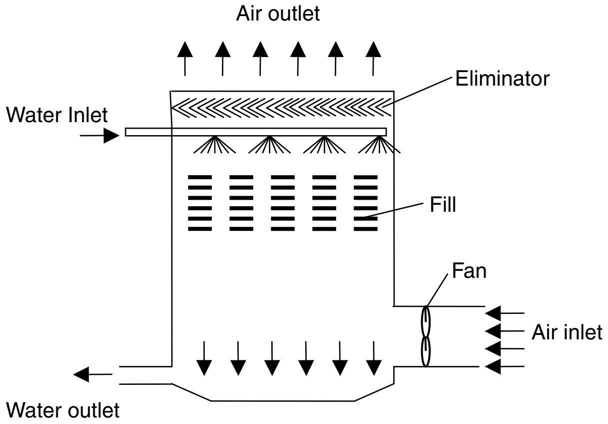


Fig. 3 Mechanical-draft counter-flow tower.

the air and evaporation of some of the water) to the atmosphere. Operation of the closed cooling towers is very similar to the open cooling tower, therefore, with one exception: the process fluid being cooled is contained in a "closed" circuit and is not exposed directly to the atmosphere or the recirculated external water.

Closed systems offer the advantages of precise temperature control (which is critical in many process applications) and low treatment cost. Because a secondary cooling system and heat exchangers are needed to cool the closed system, higher capital and operating costs are disadvantages of this design.

Mechanical and Natural Draft

In wet cooling towers, there are two types, based on the mechanism by which air is being circulated: mechanical draft and natural draft. The mechanical draft uses fans (one or more) to move large quantities of air through the tower. The mechanical draft is again divided into two types, based on the location of the air fan: forced and induced. In the case of the forced draft, the fan is located at the air entry at the base of the tower; in the induced draft, the fan is located at the air exit at the top of the tower. The induced draft produces more uniform airflow, which enhances its effectiveness over the forced draft and reduces the possibility of exhaust air recirculation.

There are many configurations of mechanical draft cooling towers that depend on the way the two fluids flow with respect to each other, such as counter flow, cross flow, and mixed flow. In a counter-flow cooling tower, air travels upward through the fill opposite to the downward motion of the water (Figs. 3 and 4).

In a cross-flow cooling tower, air moves horizontally through the fill as the water moves downward (Fig. 5 and 6). In a mixed-flow tower, air moves in a direction that is a combination of a counter flow and a cross flow. Cross-flow towers have greater air intake area, which results in considerably lower towers. This means that they have low

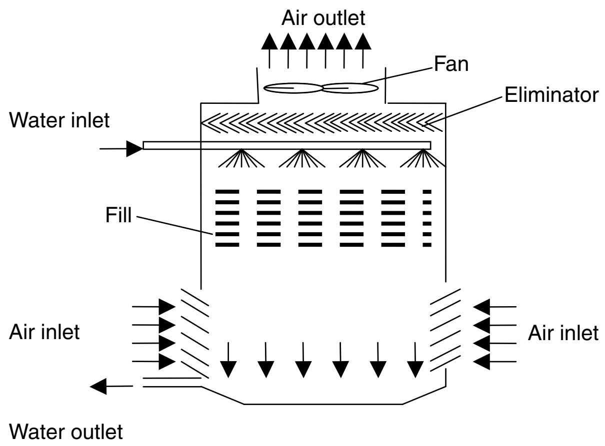


Fig. 4 Induced-draft counter-flow tower.

pressure drop in relation to their capacity and lower fan power requirement, leading to lower energy costs, but the risk of recirculation increases in tower exhaust air. On the other hand, counter-flow arrangements occupy less floor space than cross-flow towers but are taller for a given capacity, so they require higher pump heads. It shows better tower performance, since, the driest air contacts the coldest water, producing higher driving force to the heat.

A natural-draft tower is a large chimney and typically has a hyperbolic profile, which is chosen for its structural capability of withstanding wind-induced stresses and vibration; also, it requires less material. The design creates a chimney effect that causes air to move by natural convection through the fill region, which is located inside the base of the chimney (Fig. 7). As the air gets warmer from the contact with the cooling water, it gets lighter; buoyancy forces drive the air to the top of the tower and into the atmosphere, and draw fresh air into the bottom of the tower. The major economical advantage of natural-draft cooling towers is the extremely low auxiliary power consumption. Because there are no rotating parts, operational safety and low maintenance costs are maintained. The great distance between air inlet and air exit in the cooling tower prevents any hot-air recirculation

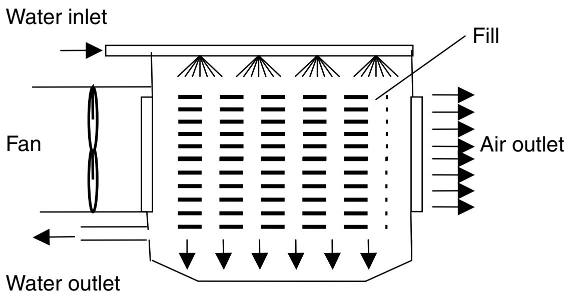


Fig. 5 Mechanical-draft cross-flow tower.

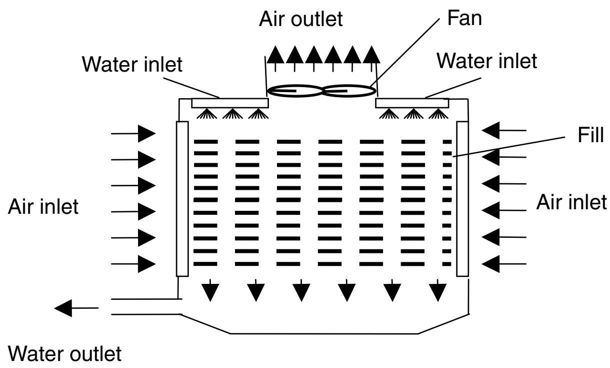


Fig. 6 Induced-draft cross-flow tower.

back to the chimney that would otherwise reduce the performance of the cooling tower. The immediate vicinity is not affected by plumes from the cooling tower, because the hot-air exit is situated at a very high elevation. The only drawback to natural-draft towers is that they are large.

There is also the assisted-draft tower, which is a natural-draft tower with some fans added at the air entry that help reduce the size of the tower. Natural-draft towers are also divided into counter flow and cross flow, defined in a similar fashion to the mechanical-draft towers.

Heat Exchange in Wet Cooling Towers

The type of heat rejection in a wet cooling tower is termed evaporative, in that it allows a small portion of the water being cooled to evaporate into a moving air stream to provide significant cooling to the rest of that water stream. The heat from the water stream transferred to the air stream raises the air's temperature and its relative humidity to 100%, and then this air is discharged to the atmosphere. The ambient air wet-bulb temperature is the controlling factor in recirculated systems and will determine the steam condensate temperature. Evaporative-heat rejection devices such as cooling towers are commonly used to provide significantly lower water temperatures than are achievable with air-cooled or "dry" heat rejection devices. The evaporative process enhances the performance of wet cooling towers over dry cooling towers severalfold due to the change in both sensible and latent heats.

Consequences and Concerns for Wet Cooling Towers

Wet cooling towers are the most common type due to their high effectiveness, but there are some drawbacks. If cooled water is returned from the cooling tower to be reused, as in the circulating systems, some water must be added to replace, or make up, the amount of the water that evaporates. Because evaporation consists of pure water, the concentration of dissolved minerals and other solids in

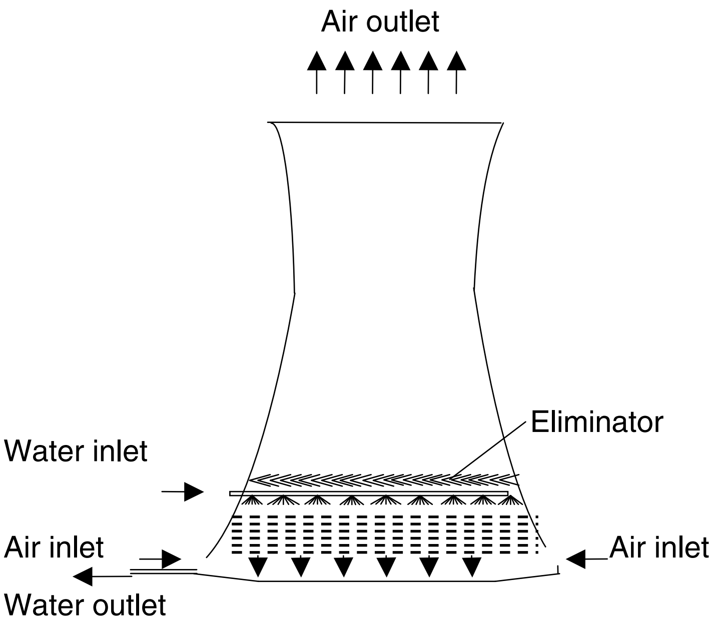


Fig. 7 Natural-draft cooling tower.

circulating water will tend to increase unless some means of dissolved-solids control (such as blow-down) is provided. Blow-down is the amount of the circulating water that is removed to maintain the quantity of dissolved solids and other impurities at an acceptable level. Some water is also lost from droplets being carried out with the exhaust air (drift). The makeup amount must equal the total of the evaporation, blow-down, drift, and other water losses (such as wind blow-out and leakage) to maintain a steady water level. Devices such as wind screens, louvers, splash deflectors, and water diverters are used to limit these losses.

The magnitude of drift loss is influenced by the number and size of droplets produced within the cooling tower, which in turn are determined by the fill design, the air and water patterns, and other interrelated factors. Drift is typically reduced by installing bafflelike devices, called drift eliminators, through which the air must travel after leaving the fill and spray zones of the tower to collect the droplets. Tower maintenance and operation levels can also influence the formation of drift, such as excessive water flow, excessive airflow, or bypassing the tower drift eliminators can increase drift emissions. Types of drift-eliminator configurations include herringbone (blade-type), wave form, and cellular (or honeycomb). The cellular units generally are the most efficient. Drift eliminators are made of various materials, such as ceramics, fibre-reinforced cement, fibreglass, metal, plastic, and wood.

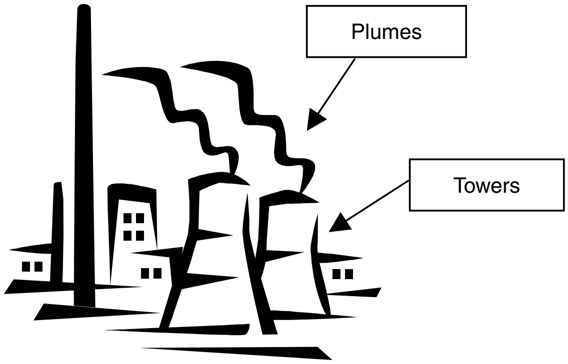


Fig. 8 Visual plume from cooling towers.

Other unfavourable environmental impacts of wet cooling are pollutant discharge—e.g., zinc, chlorine, and chromium (chromium is used to protect cooling-system equipment from corrosion)—to the atmosphere. The spread of Legionnaires' disease is due to the bacteria that thrive at temperatures typical in wet cooling systems and that can be transported through air aerosols formed in cooling towers. Other impacts are mineral drift and the formation of visual plumes. Under certain conditions, a cooling-tower plume may present fogging or icing hazards to its surroundings (Fig. 8). Some interesting pictures of cooling towers can be found at The Virtual Nuclear Tourist: Nuclear Power Plants Around the World (www.nucleartourist.com/systems/ct.htm).

Types of Fill (Packing)

The fill may consist of multiple vertical, wetted surfaces upon which a thin film of water spreads (film fill); several levels of horizontal splash elements, which create a cascade of many small droplets that have a large combined surface area (splash fill); or trickle, which is a combination of the film- and splash-type fills. A wide variety of materials and geometries have been used for packing, such as corrugated roofing sheets made of cement-based or plastic material, timber laths of triangular or rectangular cross section, plastic-impregnated paper honeycomb, and complex cellular geometries made of thin plastic material.

Thermo-Fluid Dynamic Efficiency in Cooling Towers

To choose the most convenient fill, you need to find the one that produces the maximum heat transfer with the minimum pressure drop. Other factors to be considered are the physical and chemical characteristics required for the water to cool, fouling properties, suspended materials, etc. Sirena[1] has suggested a Thermo-Fluid Dynamic Efficiency that can be used to compare and select a fill material for a particular application. This efficiency is defined as the ratio of the number of units of diffusion to the pressure drop coefficient. In this paper, the pressure drop coefficient is given for some commercial fill materials. Al-Nimr[2] has studied the dynamic thermal behaviour of cooling towers containing packing material and was able to predict closed form solutions for the transient and steady performance of a counter-flow cooling tower.

Dry Cooling

Dry cooling towers transfer heat to the atmosphere without the evaporative loss of water. Dry cooling is capable of only smaller temperature variations (around 10°C), unlike wet cooling. Similar to wet cooling, there are two types of dry cooling: direct and indirect. Direct dry cooling systems utilize air directly to condense steam, which is exhausted from the turbine into ducts and headers for distribution into rows of small-diameter finned tubes (i.e., heat exchanger). Indirect dry cooling, which is also known as the Heller System, utilises a closed-cycle water cooling system as the primary coolant to condense steam, and the heated water is air cooled. The cooling water flows through bundles of tubes with air flowing over them, but the cooling air never comes into contact with the cooling water. In both systems, direct and indirect, the flow rate of air required to achieve the same cooling capacity will have to be three or more times greater than in a wet cooling tower, so the tower will have to be much larger and more expensive.

Cooling towers are also characterized by the means by which air is moved. Similar to wet cooling towers, dry cooling towers can be mechanical draft, natural draft, or fan-assisted draft. Mechanical-draft towers rely on power-driven fans to draw or force the air through the tower. Natural-draft cooling towers use the buoyancy of the exhaust air rising in a tall chimney to provide the draft. A fan-assisted natural-draft cooling tower employs mechanical draft to augment the buoyancy effect. Many early cooling towers relied only on prevailing wind to generate the draft of air.

In power generation applications, the heat transfer between air and cooling water is achieved by convection, and the driving force of this cooling process is the approach temperature (which is defined as cooling water temperature at outlet and air temperature at inlet)—not by evaporation, as in a wet cooling tower. Some of the advantages of dry cooling towers are that they do not need any makeup cooling water or water treatment and they do not generate plume, fog, mineral drift, and disposal issues associated with wet cooling. Size is a concern in dry cooling systems, however, because dry cooling is not as effective as wet cooling; the towers have to be much larger to achieve comparable heat rejection. Another challenge to direct dry cooling is operational control in regard to how to balance the steam flow to keep the desired steam condensation temperature (and, hence, turbine back-pressure), which varies with loading. Other disadvantages of dry cooling are increase in noise, plume recirculation, maintenance of many components, and energy penalties caused by the variations of daily temperature and increases in air emissions.

Dry-cooling performance depends on the ambient air dry-bulb temperature (i.e., the sensible heat) instead of the wet-bulb temperature in the case of a wet cooling tower. Dry cooling becomes more economical when the approach temperature becomes considerably high. Other factors that affect the performance of dry cooling towers are the crosswind speed and the way that the heat exchanger bundles are arranged. The effect of crosswinds at different speeds and the effect of adding windbreak walls on the thermal performance of natural-draft dry cooling towers is given by Al-Waked and Behnia.[3] The effect of arranging the heat exchanger bundles—either vertically around the circumference of the tower or horizontally in the inlet cross section of the tower—is given by du Preez and Kröger.[4]

Wet–Dry Cooling (Hybrid)

Combined wet–dry cooling towers were introduced due to the recent stringent environmental protection laws. These towers effectively suppress detrimental plume formation at an efficiency level comparable to that of wet cooling towers. In hybrid wet–dry systems, the hot water from the power station condenser is cooled to the design discharge temperature as it passes in series first through the dry section and then through the wet section of the tower. The low-humidity hot air stream from the dry system is mixed with the moist warm air, leaving the tower at humidity levels sufficiently low to prevent the formation of visible plumes. The wet and dry components can be used separately or simultaneously for either water conservation or plume abatement purposes. At low ambient temperatures, the cooling tower can be operated as a dry cooling tower only, whereas at high temperatures, it can be used as a wet cooling tower only to achieve the required cooling without the risk of plume formation; the dry air is not put into operation until the ambient temperature starts to fall.

The design and construction of hybrid cooling towers are more complicated, and according to Streng,[5] the following data need to be specified for winter and summer operation. These data are thermal performance; cooling water flow or cooling range, which is the difference between the water temperature at inlet and the water temperature at outlet; ambient temperature; criteria for operating without plume; sound attenuation regulations; and limitations with respect to the erection area or overall height and operating weight and water analysis of the makeup water. In his work,[5] the construction, including material selection and automatic operation of the cooling system, is discussed in detail.

When a combination of wet and dry cooling technology is used, depending on system configuration, water consumption can approach that of recirculating wet systems or can be much lower. Design studies have ranged from 30 to 98% reduction in water use compared with all wet recirculating systems. As the hybrid cooling towers conform well to the stringent environmental protection requirements and to the standard operation reliability set for cooling systems, it is expected, therefore, that they will become more widespread.[5]

PERFORMANCE AND RATING OF COOLING TOWERS

In power generation, lower turbine backpressures are achieved when steam condensate temperatures are lower. Designing and operating a cooling system that can remove the heat of condensation consistently and continually at those low temperatures is essential. Therefore, the cooling system should be considered to be an integral part of the power generation process that can have a major influence on overall power plant performance and availability.

The choice of an appropriate cooling tower for a special application depends on many factors, such as capacity, availability, reliability, cost, and effectiveness. The effectiveness of a cooling tower is defined as the ratio of the actual energy that is exchanged to the maximum energy that could possibly be exchanged. The number of transfer units (NTU) is another parameter that measures the heat transfer size of the cooling tower. The higher the NTU value, the closer the cooling tower is to its thermodynamic limit. To estimate cooling-tower effectiveness and NTU, different analyses are used (similar to the ones used for analysing heat exchangers) that depend on the particular type under consideration.

The thermodynamic performance of any wet cooling tower is a function of the geometry and the ratio of the water flow rate (L) to the gas flow rate (G)—i.e., L/G. This value is quantified by means of a parameter known as the tower characteristic or number of diffusion units η ($\eta = K\,a\,V/L$), where K is the average mass transfer coefficient of condensed steam, a is the area of transfer surface per unit volume, and V is the effective volume. Manufactures supply charts for their cooling towers that present the tower characteristics as a function of L/G and the difference between the (outlet cooling water temperature (CWT) and air wet-bulb temperature (WBT).

Cooling-tower performance can be specified from the following parameters: water mass flow rate, inlet and exit temperature of water, and atmospheric wet- and dry-bulb temperatures of air. Many researchers have attempted to analyze wet cooling systems to estimate their performance. A basic theory of wet-cooling-tower operation was first proposed by Walker.[6] The practical use of basic differential equations, however, was first presented by Merkel,[7] who combined the equations for heat and water vapor transfer. He showed the utility of total heat or enthalpy difference as a driving force to allow for both sensible and latent heats. The basic approximations in Merkel's theory are

- The resistance for heat transfer in the liquid film is negligible.
- The mass flow rate of water per unit of cross-sectional area of the tower is constant (i.e., there is no loss of water due to evaporation).
- The specific heat of the air-stream mixture at constant pressure is the same as that of the dry air.
- The Lewis number (which relates heat transfer to mass transfer) for humid air is unity.
- The air exiting the tower is saturated with water vapor.

It is important to note that the formulation and implementation of Merkel's theory in cooling-tower design and rating are presented and discussed in most textbooks on unit operations and process heat transfer.

A summary of some of the methods that attempt to evaluate wet cooling towers' performance has been published by Kloppers and Kröger.[8] They compared cooling-tower performance obtained by Merkel, Poppe, and e-NTU methods. Merkel applied the mass and energy conservation laws to a differential control volume that includes the air and the water in a counter-flow cooling-tower arrangement, and derived the following differential relationships:

$$\frac{\mathrm{d}h_\mathrm{a}}{\mathrm{d}z} = \frac{h_\mathrm{D} a_\mathrm{fi} A}{\dot{m}_\mathrm{a}} = (h_{\mathrm{as,w}} - h_\mathrm{a})$$

$$\frac{\mathrm{d}T_\mathrm{w}}{\mathrm{d}z} = \frac{\dot{m}_\mathrm{a}}{\dot{m}_\mathrm{w}} \frac{1}{c_\mathrm{pw}} \frac{\mathrm{d}h_\mathrm{a}}{\mathrm{d}z}$$

where h, enthalpy; h_D, mass transfer coefficient; a, surface area per unit volume; A, frontal area; m, mass flow rate; z, vertical direction; c_p, specific heat at constant pressure;

Comp–Day

subscript: a, air; w, water; s, saturated; fi, fill; (w), evaluated at water temperature T_w.

After the above equations were combined and integrated over the whole length of the tower, the Merkel equation was derived:

$$Me_M = \frac{h_D a_{fi} A L_{fi}}{\dot{m}_a} = \frac{h_D a_{fi} L_{fi}}{G_w} = \int_{T_{wo}}^{T_{wi}} \frac{c_{pw} dT_w}{(h_{as,w} - h_a)}$$

where Me, Merkel number; L, length; G, mass velocity; subscript: M, according to Merkel approach; i, inlet; o, outlet.

The term on the right side is a measure of the cooling requirement whereas the term on the left side is a measure of the performance of the packing.

Poppe included the effect of Lewis factor Le (defined as $h_c/h_D\ c_{pa}$) and the reduction in water flow due to evaporation. He derived two equations for the Merkel number based on the state of the air at exit—unsaturated or supersaturated. If the air is exiting as unsaturated, the Merkel number can be obtained by an iterative procedure of integrating the following equation:

$$\frac{dMe_P}{dT_w} = c_{pw} / \Big[h_{as,w} - h_a + (Le - 1)\{h_{as,w} - h_a - (w_{s,w} - w)h_v\} - (w_{s,w} - w)c_{pw}T_w \Big]$$

In the above equation, w, humidity ratio; the subscript P, Poppe approach; v, vapor; a, air; s, saturated, and w, water.

The Merkel number for air exits as supersaturated can be obtained by an iterative procedure of integrating the following equation:

$$\frac{dMe_P}{dT_w} = c_{pw} / \Big[h_{as,w} - h_{ss} + (Le - 1)\{h_{as,w} - h_{ss} - (w_{s,w} - w_{s,a})h_v\} - (w - w_{s,a})c_{pw}T_w \Big] + (w - w_{s,w})c_{pw}T_w$$

The subscript ss = supersaturated. Details can be found in Poppe and Rögener[9] and in Bourillot.[10]

According to the e-NTU method, in which the same simplification of Merkel is used, the Merkel number for the case where dry air mass flow rate $m_a > m_w\ c_{pw}/(dh_{asw}/dT_w)$ can be obtained by:

$$Me_e = \frac{c_{pw}}{dh_{as,w}/dT_w} \text{NTU}$$

The subscript e = the e-NTU approach.

The Merkel number for the case where dry air mass flow rate $m_a < m_w\ c_{pw}/(dh_{asw}/dT_w)$ can be obtained by:

$$Me_e = \frac{\dot{m}_a \text{NTU}}{\dot{m}_w},$$

where NTU is given by:

$$\text{NTU} = \frac{1}{1 - C} \ln \frac{1 - e^C}{1 - e}.$$

C is the fluid capacity rate ratio and is defined as C_{min}/C_{max}. It is to be noted that the e-NTU method is applicable to the cross-flow arrangements, provided that the air and water streams should be defined, whether they are mixed, unmixed, or a combination.

Khan and Zubair[11] presented an analysis to estimate the effectiveness and NTU of a counter-flow wet cooling tower that matched the experimental data closely. They included in their model the effect of the Lewis number, defined in a similar fashion to Poppe's as the ratio of the convective heat transfer coefficient to the convective mass transfer coefficient times the specific heat at constant pressure of moist air, the heat resistance in the air-water interface, and the effect of water evaporation on the air states along the vertical length of the tower. In their analysis, they assumed constant convective heat and mass transfer coefficient, and ignored the heat lost through the tower walls, variation in specific heat properties, and water lost by drift. They applied the mass and energy conservation equations to a differential volume to relate the change in enthalpy of moist air to its humidity ratio, in terms of Lewis number Le and other properties of moist and saturated air. The outlet properties are obtained by numerically integrating the set of differential equations of conservation of mass and energy on an increment volume of the cooling tower.

They also gave the following definition of NTU and effectiveness:

$$\text{NTU} = \frac{h_D A_V V}{\dot{m}_a} = \int_{W_i}^{W_o} \frac{dw}{w_{s,w} - w}$$

$$\varepsilon = \frac{h_o - h_i}{h_{s,w,i} - h_i}.$$

A_V, surface area of water droplet per unit volume of the tower; V, tower volume; subscript: o, outlet, and i, inlet.

They also gave an empirical equation for NTU_{em}:

$$\text{NTU}_{em} = c\left(\frac{\dot{m}_w}{\dot{m}_a}\right)^{n+1}$$

c and n, empirical constants specific to a particular tower design, and subscript em, empirical.

Another approach is to return to the fundamental equations of fluid mechanics and heat and mass transfer, and arrive at numerical solutions with the aid of computational fluid dynamics technique (CFD). Some examples are the work by Al-Waked et al.[12] and Hasan et al.[13] among many others. These solutions can in principle be used as the sole basis of design or they can be used to examine, modify, and improve existing simpler methods—such as work by Kloppers and Kröger,[14] who used the finite difference method to compare the three approaches of Merkel, Poppe, and e-NTU.

CONCLUSION

The different types of cooling towers—wet, dry, and hybrid—have been presented. Research and experience show that the hybrid cooling towers conform well to the stringent environmental protection requirements and to the standard operation reliability set for cooling systems; it is expected, therefore, that they will become more widespread. Different methods to estimate cooling-tower performance are presented, based on some assumptions that simplify the problem. As the systems get more complicated, however, CFD is capable of predicting performance and can be used as the sole basis of design, or it can be used to modify and improve existing simpler methods to make them closer to reality.

Glossary

Approach temperature: The difference between the temperature of the condenser water leaving the tower and the wet-bulb temperate of the air entering the tower in the case of the wet tower and the dry-bulb temperature of the air entering the tower in the case of the dry tower.

Blow-down: The quantity of the circulating water that is removed to maintain the amount of dissolved solids and other impurities at an acceptable level.

Blow-out: Water droplets blown out of the cooling tower by wind—generally, at the air inlet openings. In the absence of wind, water may also be lost through splashing or misting.

Drift: Water droplets that are carried out of the cooling tower with the exhaust air.

Drift eliminator: Equipment containing a complex system of baffles designed to remove water droplets from cooling-tower air passing through it.

Noise: The sound generated by the impact of falling water; the movement of air by fans; the fan blades moving in the structure; and the motors, gearboxes, and drive belts.

Plume: The stream of saturated exhaust air leaving the cooling tower. The plume is visible when the water vapor it contains condenses in contact with cooler ambient air.

Range: The difference between the cooling water temperature entering the tower and the cooling water temperature leaving the tower.

More glossary words are available at the Cooling Technology Institute (CTI) Web site (www.cti.org/whatis/coolingtowerdetail.shtml).

REFERENCES

1. Sirena, J.A. The use of a thermo-fluid dynamic efficiency in cooling towers. Heat Trans. Eng. **2002**, *23* (2), 22–30.
2. Al-Nimr, M.A. Modelling the dynamic thermal behaviour of cooling towers containing packing materials. Heat Trans. Eng. **1999**, *20* (1), 91–96.
3. Al-Waked, R.; Behnia, M. The effect of windbreak walls on the thermal performance of natural draft dry cooling towers. Heat Trans. Eng. **2005**, *26* (8), 50–62.
4. du Preez, A.F.; Kröger, D.G. Effect of the heat exchanger arrangement and wind-break walls on the performance of natural draft dry-cooling towers subjected to cross-winds. J. Wind Eng. Ind. Aero. **1995**, *58* (3), 293–303.
5. Streng, A. Combined wet/dry cooling towers of cell-type construction. J. Energ. Eng. **1998**, *124* (3), 104–121.
6. Walker, W.H.; Lewis, W.K.; McAdams, W.H.,; Gilliland, E.R. *Principles of Chemical Engineering*; McGraw-Hill, Inc.: New York, NY, 1923.
7. Merkel, F. Verdunstungshuhlung. Zeitschrift des Vereines Deutscher Ingenieure (V.D.I) **1925**, *70*, 123–128.
8. Kloppers, J.C.; Kröger, D.G. Cooling tower performance evaluation: Merkel, Poppe, and e-NTU methods of analysis. J. Eng. Gas Turbines Power **2005**, *127* (1), 1–7.
9. Poppe, M.; Rögener, H. *Berechnung von Rückkühlwerken, VDI-Wärmeatlas*; 6. Auflage, VDI Verlag GmbH : Berlin, Germany, 1991.
10. Bourillot, C. TEFERI: Numerical Model for Calculating the Performance of an Evaporative Cooling Tower. EPRI Report CS-3212-SR, Electric Power Research Institute: Palo Alto, CA, 1983.
11. Khan, J-U-R.; Zubair, S.M. An improved design and rating analyses of counter flow wet cooling towers. J. Heat Trans. **2001**, *123* (4), 770–778.
12. Al-Waked, R.; Behnia, M. The performance of natural draft dry cooling towers under crosswind: CFD study. Int. J. Energ. Res. **2004**, *28* (2), 147–161.
13. Hasan, A.; Guohui, G. Simplification of analytical models and incorporation with CFD for the performance prediction of closed wet cooling towers. Int. J. Energ. Res. **2002**, *26*, 1161–1174.
14. Kloppers, J.C.; Kröger, D.G. A critical investigation into the heat and mass transfer analysis of counter flow wet-cooling towers. Int. J. Heat Mass Trans. **2005**, *48* (3–4), 765–777.

BIBLIOGRAPHY

Some useful sites and references that were used to collect some of the information in this article are:

1. Hewitt, G.F., Shires, G.L., Polezhaev, Y.V., Eds., *International Encyclopaedia of Heat and Mass Transfer*, CRC Press: New York, NY, 1997.
2. Shan, K.W. *Handbook of Air Conditioning and Refrigeration*, 2nd Ed.; McGraw-Hill, Inc.: New York, NY, 2001.
3. Cooling Technology Institute. Available at: www.cti.org.
4. Gulf Coast Chemical Commercial, Inc. Available at: www.gc3.com/techdb/manual/coolfs.htm

Comp–Day

5. U.S. Environmental Protection Agency. Available at: www.epa.gov/waterscience/316b/technical/ch4.pdf#search=‘wetdry%20cooling%20towers
6. Online Chemical Engineering Information. Available at: www.cheresources.com/ctowerszz.shtml
7. GEA Cooling Tower Technologies. Available at: www.bgrcorp.com/default-gct.htm
8. Lenntech. Cooling Towers. Available at: www.lenntech.com/cooling%20towers.htm
9. Wikipedia. Available at: www.en.wikipedia.org/wiki/Cooling_tower
10. Legionella Control. Available at: www.legionellacontrol.com/legionella-glossary.htm

Comp–Day

Data Collection: Preparing Energy Managers and Technicians☆

Athula Kulatunga
Department of Electrical and Computer Engineering Technology, Purdue University, West Lafayette, Indiana, U.S.A.

Abstract
Energy audits can be used to provide hands-on activities related to an energy management course. After learning the necessary background concepts, students need to be aware of what measurements must be taken to evaluate an existing energy system. In industry and universities, one may find apprentices and students from different educational backgrounds, such as electricians with no exposure to newer measuring instruments and students with no ability to take electrical and/or mechanical measurements. By studying the capabilities and limitations of measuring instruments, newcomers to the energy auditing may collect reliable data. This article introduces several hands-on activities that could be replicated to teach students how to take accurate measurements of electrical, light, and heat flow parameters, ultrasonic leak detection, electronic combustion analysis, and simple data acquisition before conducting energy audits. A sample laboratory activity includes a description of the measuring instrument, factors that contribute to inaccurate readings, safety concerns, and several practice measurements useful to energy audits.

INTRODUCTION

It has been accepted that engineering technology courses should have some hands-on activities such as labs, projects, and other practical experiences. In the field of energy management, energy audits have been used effectively to provide hands-on experiences.[1] An energy audit, also known as energy survey, energy analysis, or energy evaluation, is a process that examines the current energy consumption of a process or facility and proposes alternative ways to cut down energy consumption or costs. One aspect of the energy auditing process is to collect specific data of a process or a facility. Measuring temperature, flow rates (heat, liquid, and air), intensity of light, electrical current, voltage, power, power factor (PF), humidity, pressure, or vibration may be required to determine the energy consumption and waste. New measuring equipment is pouring into the measurement world making data collection easier, more accurate, and safer.

Accurate data collection is paramount not only to analyze energy consumption, but also to evaluate the effectiveness of proposed changes suggested in an energy audit report. Some energy-saving electrical retrofits may introduce electrical power quality problems that may not be accounted for by traditional meters, causing erroneous data. With inaccurate data, the conclusion of an energy improvement project holds no validity.[2]

Preliminary or walk-through energy audits are the most suitable for beginners. A preliminary energy audit is a process during which an auditor examines an existing energy consuming system according to a predetermined set of procedures. The procedures are outlined as a result of a historical data analysis of the targeted system and conversations the auditor had with the owner or the operator of the system. These procedures include taking electrical and other measurements under certain conditions.

ELECTRICAL MEASUREMENTS

Data collection of any system that consumes electrical energy requires at least three basic measurements—voltage, current, and PF—for energy analysis calculations. Utility meters collect all these data at a building service entrance point or at any other sub-metering location, if such meters are installed. By contacting the utility provider, one can easily obtain the historical data related to the above parameters and more for a given facility. However, when it comes to individual systems within a facility, these data may not be available, unless additional utility meters are installed at each of the service entrance points to individual systems. In a modern power distribution system, harmonics of the fundamental frequency, 50 or 60 Hz, appear due to non-linear devices connected into the system. Power line harmonics cause erroneous readings, if a meter is not capable of measuring

☆ This entry originally appeared as "Preparing Energy Managers and Technicians for Energy Data Collection" in *Energy Engineering*, Vol. 102, No. 4, 2005. Reprinted with permission from AEE/Fairmont Press.

Keywords: Energy audits; Energy data; Energy measurements.

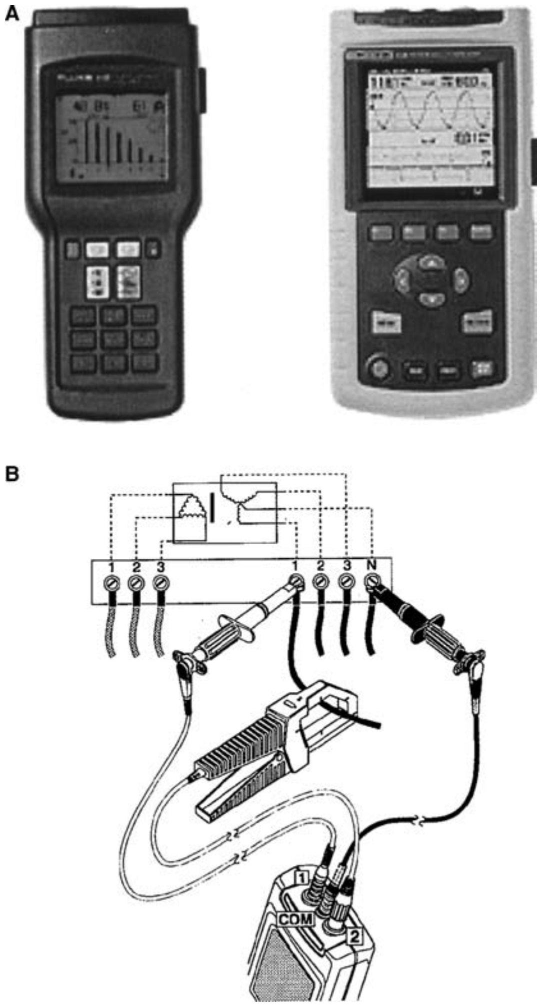


Fig. 1 (A) Fluke 41B and 43B power analyzer, (B) Using meter probes for three-phase. (Courtesy of Fluke Corporation).

true root-mean-square (RMS) values of the fundamental frequency and the harmonics at a given instant.[3]

Two power quality analyzers made by Fluke Corporation are shown in Fig. 1A. One of these meters can be easily used to obtain all necessary data for an electrical system. Fig. 1B shows how to obtain measurements for a three-phase balance (5% or less imbalance) using voltage and current probes that come with the meter. The meters are capable of measuring true RMS, peak, and total harmonic distortion (THD) for voltage and current. They also display true power (Watts), reactive power volt ampere reactance (VAR), PF, displacement power factor (DPF), crest factor, *K*-factor, and harmonics individually up to the 31st.

The meters are capable of displaying data in three views—waveforms, barographs showing harmonic levels, and numeric values for voltage, current, and power—as depicted in Fig. 2.

Each of the data displayed on Fig. 2 gives some clue about the nature of power quality. For example, flat-topped voltage waveform is an indication of the presence of current harmonics. One can use the THD levels to determine if they are within the specified limits. If not, the same data can be used to determine sizing of transformers and harmonic filters. The bar-graph display of the current reveals the percentage comparison of odd-numbered harmonics with respect to the fundamental frequency. Even-numbered harmonics cancel out in a power system; therefore their effects are not a concern in a power system. Finally, the meter displays numerical values such as total PF, DPF, kilovolt ampere, and kilowatt useful in determining power-factor correction methods. The data can be further analyzed by downloading FlukeView® software, which comes with the meter. Fig. 3 depicts current waveform of a nonlinear load and downloaded data in a tabular form.

Fluke 43B power quality analyzer has all the features of a 41B as well as sample and storage capabilities. These additional features allow a user to detect voltage/current sags and transients in a power system.

Practice Activity Outline

A typical electrical power system may have one of the following problems: (1) voltage sags, (2) current balance and loading, (3) harmonics, (4) grounding, or (5) loose connections. Taking measurements at the electrical service panel during an energy audit, the investigator can determine the sources of these problems. In a laboratory environment (or in a workshop), each of the above problems can be replicated to demonstrate the effects.

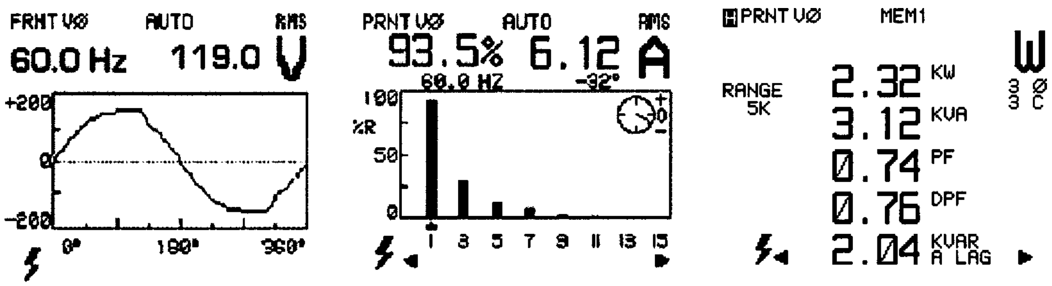


Fig. 2 Waveform, bar graph, and numerical displays of data presented by Fluke 41B. (Courtesy of Fluke Corporation).

Summary Information			Voltage	Current
Frequency	59.98	RMS	116.26	0.85
Power		Peak	162.02	2.46
Watts	61.00	DC Offset	0.05	-0.04
VA	98.00	Crest	1.39	2.90
Vars	12.00	THD Rms	1.93	77.09
Peak W	405.00	THD Fund	1.94	121.03
Phase	11° lead	HRMS	2.25	0.65
Total PF	0.62	KFactor		15.81
DPF	0.98			

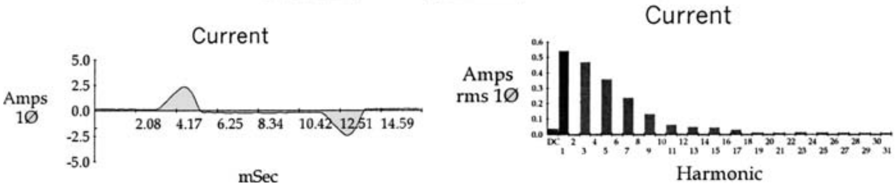


Fig. 3 Data table, current waveform, and harmonic representation of a non-linear load taken by Fluke View® software.

For example, the circuit shown in Fig. 4 may help students learn how to detect harmonics and locate the harmonic producing sources in a power system. The three-phase power supply should be taken from a Y-connected transformer secondary where the common point is taken as neutral. It is also necessary to ground the common point, if concepts related to overloading are introduced. A double-pole double-throw (DPDT) switch allows switching between the lamp-only circuit and the lamp with a dimmer circuit. The lamp-only path would allow the student to study the waveform characteristics and harmonic content of a linear load. The lamp with a dimmer can be used to observe the voltage and current waveform variation when harmonics are present to the right of the point, where the Fluke 43B meter is connected. Using the sag and swell mode of the Fluke 43B meter, a sample can be taken when DPDT is in position A and the Variable Speed Drive (VSD) is off. The waveform would look similar to the one shown in Fig. 5A. Note that the voltage sag coincides with a current swell indicating that the disturbance has occurred downstream of the measurement point.

The concept of upstream disturbance can be demonstrated by switching the DPDT switch to position B and starting the drive. A sample display taken under this condition is shown in Fig. 5B. Note that the voltage sag occurs simultaneously with the current sag, indicating that the source of the disturbance is upstream of the measurement point.[4]

Precautions

First of all, students must have a thorough understanding of the instrument's limitations. They should make voltage measurements after a circuit breaker and wear safety gloves and eyewear when taking measurements of live panels.

LIGHT MEASUREMENTS

On average, lighting consumes 35% of energy used in commercial buildings and 25% in industrial facilities. Lighting levels directly affect the productivity of

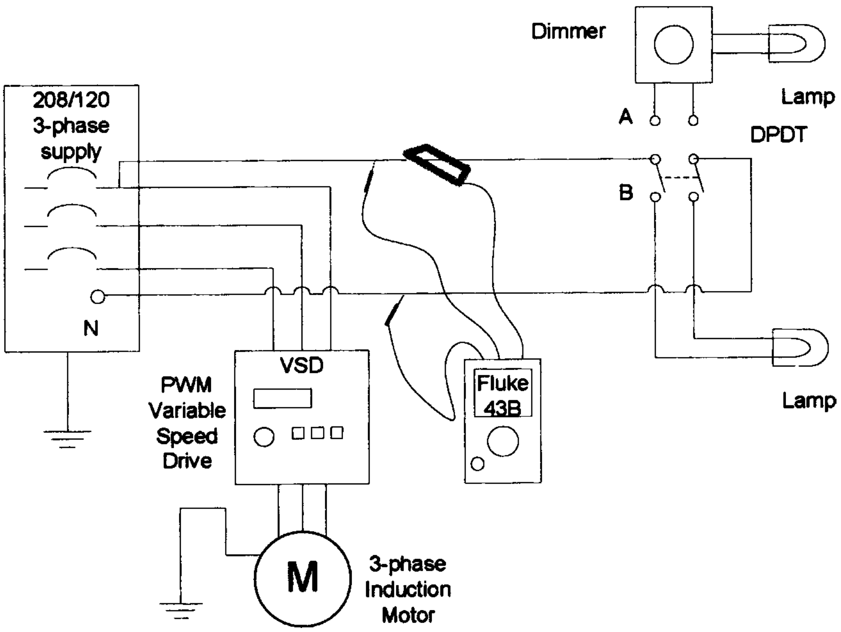


Fig. 4 A circuit for learning the sources of harmonics in an electrical power system.

Comp–Day

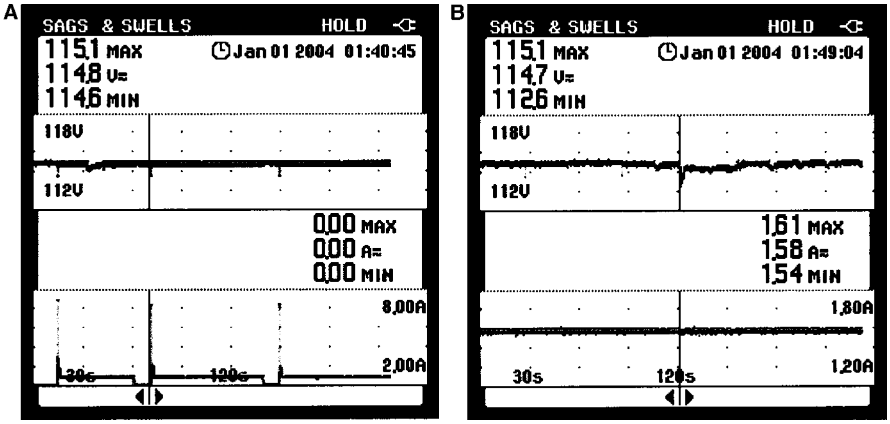


Fig. 5 (A) Downstream disturbance, (B) Upstream disturbance.

employees. However, many lighting systems are improperly designed and unattended over the years. Lighting is one of the areas where companies can save energy with the least amount of capital investments. Fig. 6 illustrates the before and after appearance of a warehouse, where 50% more light was obtained.

Light meters are very easy to use. The meter in Fig. 7 measures light intensity in Lumen and light density in foot-candles (fc). The purpose of measuring light is to determine the incident light on a horizontal or vertical surface. Illumination Engineering Society of America (IESA) publishes recommended light levels for specific tasks.

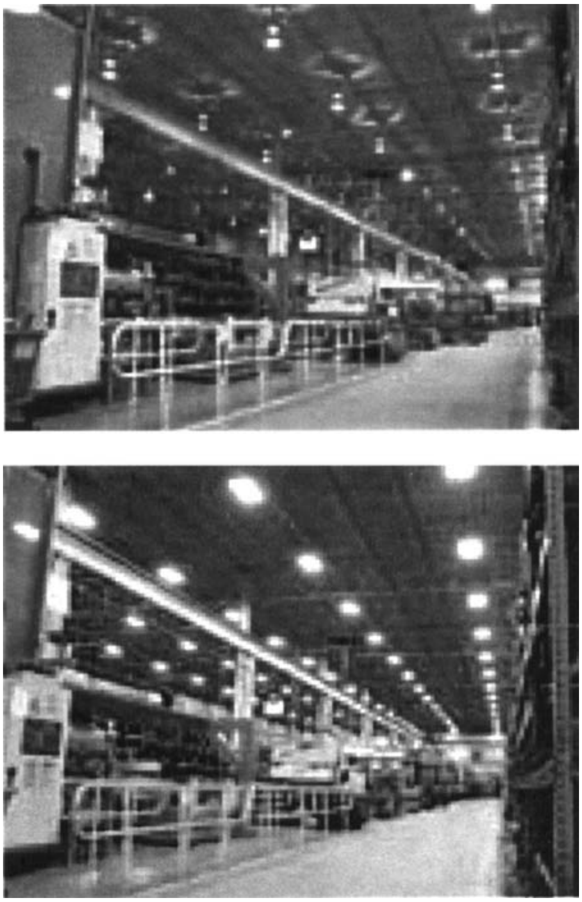


Fig. 6 Fifty percent more light obtained through lighting retrofits in a warehouse.
Source: From Business News Publishing Company (see Ref. 5).

Practice Activity Outline

Select a location where students can move around with a ladder for at least 2 h. First, students draw an accurate layout of the space being audited including the location of furniture, windows, cabinets, and lighting fixtures. The location could be a classroom, a lab, a corridor, or a gym.

Using a light meter, students record the available light intensity in fc on horizontal and vertical surfaces with and without artificial lights. This would yield a light distribution map. Now compare the available light with the recommended light to determine levels and locate poorly illuminated or over-illuminated areas. Most classroom and labs are illuminated by fluorescent light fixtures. The amount of light generated by fluorescent light deteriorates with time. Students may replace existing fluorescent tubes with new ones to examine the available light.

Precautions

Students need to be aware that only the incident light on horizontal surfaces can be added. Light intensity must be measured at the recommended height as specified by IESA or any other standard. It is important to stay away from the detector to minimize the effect of body-reflected light entering into the detector.

BUILDING ENVELOP MEASUREMENTS

The heat loss or gains of a building depend on many factors: *R*-value of walls and roof material, window characteristics, ambient temperature, humidity level, and heating and cooling degree days of a given location where

Fig. 7 A stick-type light meter with a retractable sensor. (Courtesy of Omega Inc.).

the building is allocated. When *R*-value and *U*-values are known, a detailed analysis of a building envelop would yield British thermal unit lost or gained through walls, windows, and other heating and cooling sources. It is very difficult to determine the *R*-values of even a several-year old facility due to poor record keeping and later add-ons to the structures. The OMEGA® OS-650 energy conservation and plant maintenance kit (see Fig. 8) is very useful in energy audits and general plant maintenance. The kit consists of an infrared (IR) thermometer capable of measuring temperatures from −2 to 200°F and a heat flow meter, which is a specially designed IR radiometer capable of displaying heat flow though a scanned wall in terms of British thermal unit per square feet hour.[6] The kit is priced around $1600. If purchased separately, either the thermometer or the heat flow meter would cost about $800.

Practice Activity Outline

A laboratory activity can be developed, so students will be able to (1) estimate *R*-value value of an unknown insulator, (2) measure heat flow through walls and windows, and (3) make energy cost analyzes at the end of the activity. To estimate *R*-value, net heat flow should be determined by using the heat flow meter. This is done by taking two measurements across a wall at the same height—one from inside and the other from outside of the room. The meter shows a + or − number, indicating the direction of heat flow.

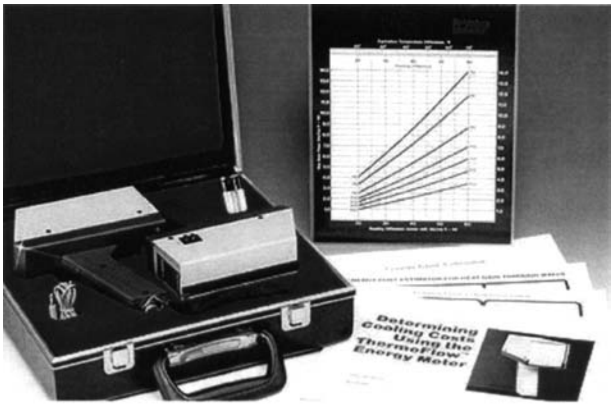


Fig. 8 OMEGA OS-650 Energy Conservation and Plant Maintenance Kit. (Courtesy of Omega Corporation).

THERMAL IMAGING

During any energy conversion process, heat is produced as a by-product. Poor insulation of buildings and pipes carrying hot or cold liquids causes energy wastes. New retrofits and devices containing switching power electronic circuits may introduce harmonics, which may overheat the neutral conductors. Poor connections of an electrical distribution system are the major contributor to system inefficiencies and may lead to a catastrophic fire. All of the above could be avoided, if one could detect them in advance. To detect such abnormalities, one requires a thermal imager. A thermal imager, compared to an IR thermometer, is capable of measuring temperature variation between two adjacent points. In the past, the cost of thermal imagers prevented widespread use in energy management and plant-maintenance activities. Industries hired a consultant to survey all electrical distribution panels and other critical locations in a facility to determine the thermal profile annually. But prices have come down significantly, under $10,000, over the past few years. A specially designed, hand-held thermal imager made for energy audits and plant maintenance is shown in Fig. 9.

The imager can hold up to 100 images, which are stamped with time and date as they are taken. Through an Universal Serial Bus (USB) port, the images can be downloaded to the accompanying software installed on a PC for further analysis. Images can be further analyzed by assigning a single color to a temperature range and creating a thermal profile. A thermal profile represents the temperature at the *x* and *y* axes as the cursor moves around the image.[7]

Practice Activity Outline

Many interesting activities can be developed around this device. Students could measure the temperature of different light sources and compare the power consumption of each light bulb and the surface temperature. The internal heat distribution of a room could be detected and documented by hanging black-painted aluminum sheets. Since this device measures the reflected thermal energy, the manufacturer recommends painting highly reflective surfaces with some dark color to minimize reading errors. Students could develop thermal images of buildings on campus and analyze the heat losses. In a power lab, students could measure the temperature of conductors when motors are driven with variable speed drives (Fig. 10).

Precautions

When taking an image, the focus is paramount. As with any digital camera, lack of focus blurs the image, thus minimizing the device's ability to distinguish the temperature difference between adjacent pixels. Some

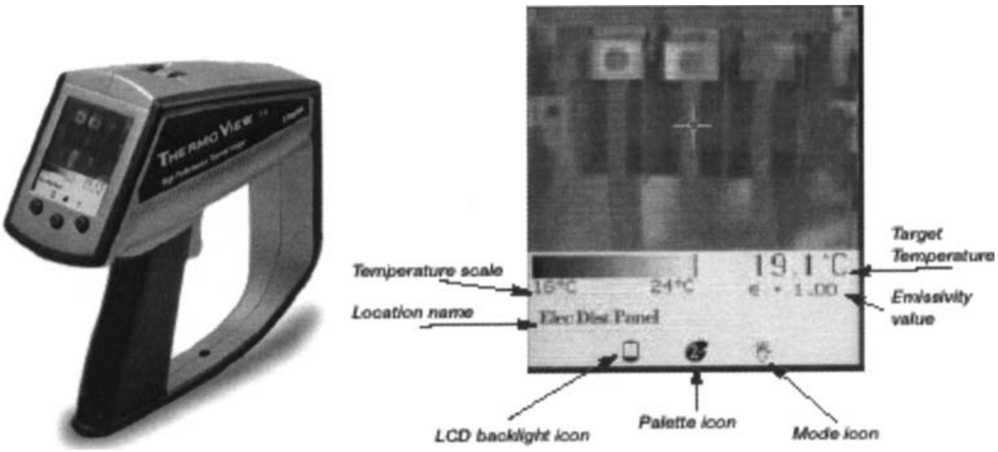


Fig. 9 Raytek ThermoView™ Ti30 portable imager and its display. (Courtesy of Raytek Corporation).

imagers have a minimum focusing distance that must be adhered to. Most current thermal imagers allow users to view the object using different color palettes such as rainbow, ironbow, and grayscale. Despite the popularity of color palettes, it is recommended to use grayscale for most applications, because the human eye can detect variations of grayscale better when thermal changes are subtle. In addition, students must be aware of concepts such as qualitative vs quantitative temperature measurements, distance to target ratio, field of view, effects of environmental conditions (steam, dust, smoke, etc.), and effects of emissivity.

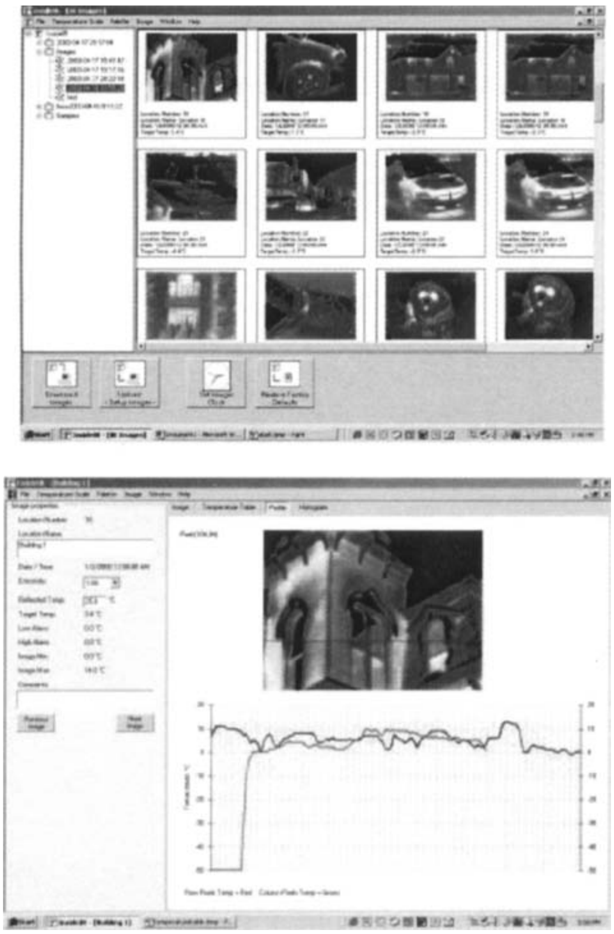


Fig. 10 Multi-image display and thermal profile using ThermoView™ Ti30 imager. (Courtesy of Raytek Corporation).

COMBUSTION ANALYSIS

The effectiveness of fossil-fuel burning receives little attention during an energy audit process unless the auditor is well trained to perform mechanical tests. The mechanical test, a tedious process, requires a set of data from several different pieces of equipment such as draft gauges, thermometers, carbon monoxide (CO) stain length tubes, wet chemical absorption instruments, etc. Then the data are entered into a slide ruler calculator to determine the combustion efficiency.[8]

New electronic combustion analyzers provide more reliable and provable data than a traditional "eyeballing the flame" analysis. It also produces faster professional analysis than a mechanical test. Fyrite® Tec 50 and 60 residential combustion analyzers by Bacharach Inc. shown below (Fig. 11) falls under $500 and comes with many features that make this instrument ideal for technicians and students alike.

A user first selects the type of fuel being burnt. The fuel choices are natural gas, #2 oil, propane, and kerosene. The meter can measure flue gas oxygen content 0.0%–20.95% O_2, flue gas temperature up to 999°F, ambient temperature 32°F–104°F, and flue gas CO content 0–2000 ppm CO. Based on these measurements, the meter is capable of calculating the following: combustion efficiency 0.1%–99.9%, flue gas carbon dioxide content 0.0 to a fuel dependent maximum (in percent), flue gas CO air-free content 0–9999 ppm, and excess air 0%–400%.

	Eyeball Analysis	Mechanical Tests	Electronic Analyzers (Fyrite Pro)
Reliability	poor	good	excellent
Accuracy	poor	good	excellent
Accurate/Safety Measurements (CO)	no	no	yes
Ease of use	n/a	no	yes
Method of Test	blind	snapshot	continuous
Speed	moderate	slow	fast
Documentation Capability	no	no	yes
Cost of Ownership (3 years)	n/a	moderate	low

Fig. 11 Fyrite® Tec 50 and 60 residential combustion analyzers and a manufacturer's comparison of combustion analysis techniques. (Courtesy of Bacharach Inc.).

Practice Activity Outline

Practice activities for this meter may include: (1) testing flames produced by the four different type flames while varying the air intake into the flame, (2) experimenting heat transfer characteristics of a jar containing water (or any other liquid applicable to a certain industrial process) under different insulations and ambient temperatures, and (3) simulating a specific type of burner used in an industrial facility to investigate efficiency improvement opportunities. Flames for activity-1 and activity-2 above can be set up by using off-the-shelf burners or commercial water heaters.

Precautions

When working with flammable gases in closed environments, one must adhere to Occupational Safety and Health Administration (OSHA) safety guidelines. Students must be aware of every detail of the gas being used and wear appropriate protection apparatus, which include burn-proof gloves and eye protection.

ULTRASONIC LEAK DETECTION

A leak, whether it is a compressed air, steam, or conditioned air, is a waste. Hidden costs due to leaks can be significant in very competitive market environments. Some leaks are very easy to detect, while others are not. Most leaks go undetected due to ambient noises. To detect some types of leaks, one has to use ultraviolet (UV) dyes, special lamps, bubble solutions, etc. Leak detection takes time and money. Any instrument that detects different types of leaks within a very short time would be the most practical solution.

Ultrasonic leak detectors minimize traditional problems associated with leak detection. Fig. 12 depicts a leak detector made by Superior Signal Company Inc., which comes with a price of less than $250.

According to the manufacturer's specifications, the meter can detect leaks of any type of refrigerant gas, vacuum, and pressure leaks. The reading is not affected by saturated gas areas or windy environments. An energy auditor may even detect leaks around freezer and cooler doors with this meter.[9]

Fig. 12 AccuTrak VPE by Superior Signal Company, Inc. (Courtesy of Superior Signal Company, Inc.).

Practice Activity Outline

Several simple activities can be developed to familiarize this equipment in a laboratory or workshop environment where compressed air is available. Several holes, at least one foot apart with different diameters, can be drilled in a copper piping tube. A connector needs to be soldered to one end and the other end must be capped. Once connected to a regulated compressed air outlet, learners may trace air leaks of each hole under different pressure to learn the nature of the leaks and the corresponding meter responses. The activity can be further expanded by moving away from the leaks and/or walking around the pipe. Similarly, to study the nature of leaks due to doors and windows, learners may take measurements around commercial soft drink coolers located in any facility.

Precautions

People have a natural tendency to clear the area around a suspected leak, especially painted or corroded surfaces, to examine the leak more closely. This may cause the hole to burst without much warning, spewing dust and loose particles all over. Therefore, students must wear safety glasses and dusk masks as appropriate when testing leaks of any form.

LOW COST DATA ACQUISITION

Energy managers have to rely on historical data when energy analyzes are performed. A trend of a measured parameter such as current, voltage, temperature, etc. presents a better picture of energy-saving opportunities than instantaneous readings. Most data loggers are expensive, because they are designed to minimize potential damages to the front-end electronics of the data acquisition system and the computers used to record the data. For laboratory applications and short-term data logging, one may use the D1-194RS low-cost (less than $25) data acquisition starter kit shown in Fig. 13. The kit includes hardware, recorder software, and a serial port cable.

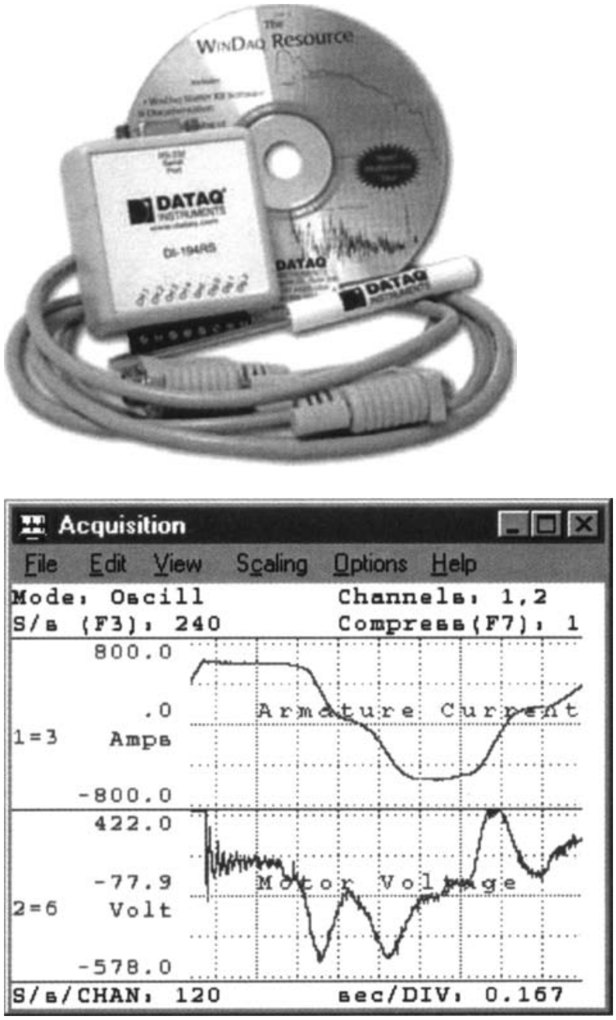


Fig. 13 DATAQ 4-channel data acquisition and chart recorder starter kit model. (Courtesy of DATAQ Instruments, Inc.).

The hardware is self-powered through the serial port of PC. Analog signals connected to any of the four channels are digitized and saved into the hard drive while showing on the screen as a strip-chart recorder. Each channel can be sampled up to 240 samples per second with 12-bit resolution, which is adequate for most analog signals in the energy field. Each channel accepts ± 10 V, which is large enough for most commercially available transducer outputs. The software includes Active-X control libraries that allow the user to program the kit from any Windows environment. The recoded data may be play backed for later analysis.[10] The input signals are not optically isolated and measured with respect to a common ground point.

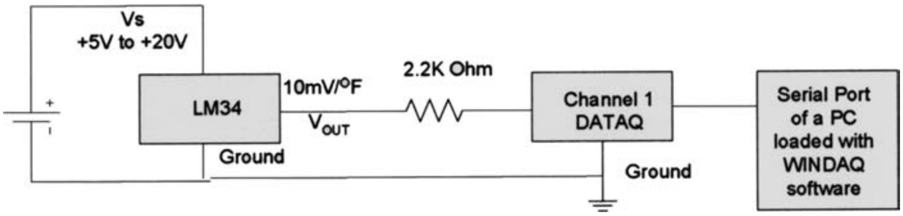


Fig. 14 Schematic diagram to measure ambient temperature directly with DATAQ starter kit.

Practice Activity Outline

Any transducer that produces 0 to +10 V or 0 to −10 V output proportional to the measured variable can be connected directly. As a precaution, measure the voltage between two grounds to verify that the two grounds are at the same electrical potential. When commercial transducers are not available, a very inexpensive ambient room temperature sensor could be developed by using a LM34 temperature sensor integrated circuits (IC) made by National Semiconductors Inc., as shown in Fig. 14.

LM34 IC produces 10 mV for every degree of Fahrenheit. For example, when it senses 72°F, the output would be 720 mV. The computer display would show this reading as a 720 mV, or it can be calibrated on the software to display in Fahrenheit. The software can be set up to save time-stamped signals for later analysis. The activity can be expanded by adding a second channel that tracks the turn-on–off signal of rooms' heating ventilating and air conditioning (HVAC) unit. Once the graphs are plotted, the learners may be able to study process characteristics of the room by measuring dead time, time constant, etc.

Precautions

Users of this data acquisition system should be aware that this unit does not provide electrical isolation. As long as the transducer's electrical ground and the unit hardware ground, which is the PC ground, are at the same potential, the unit will work accurately. A dc voltmeter reading between these two points reveals any potential problems. The unit measures only voltage signals. Current signals should be converted into a proportional voltage signal using a current shunt placed in series with the load. However, common mode voltage applied to the unit may destroy the unit hardware unless necessary precautions are taken. There are many useful literatures available in www.dataq.com website that would help users in making measurements safe and easy.

CONCLUSIONS

New meters appearing on the market make data collection for energy audits fun and instructive. With this new measuring equipment, an auditor can take measurements that were not economical or even possible several years ago. Collected data can be easily analyzed with software and sent to another individual conveniently over the internet. Seven measurement instruments have been presented. Each meter has its own unique capabilities compared to the traditional equipment. However, they have their own limitations as well. When taking sophisticated measurements, one should be aware of the conditions under which those measurements must be taken. One can understand the measuring equipment potentials and limitations by performing a set of controlled activities in a laboratory environment. This will allow the learner to change one variable at a time. Under no circumstances should one bypass safety requirements specified by the equipment manufacturers. Some meters require periodical calibrations.

REFERENCES

1. Kulatunga, A. *Integrating the Courses in Electrical and Mechanical Engineering Technologies to Fulfill the Needs of Energy Professionals*, Proceedings of ASEE Conference, Montreal, Canada, June 16–19, 2002.
2. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 3rd Ed.; The Fairmont Press: Lilburn, GA, 2000; [ISBN0-88173-336-9].
3. Kulatunga, A. *Peak Current Behavior of Nonlinear Loads in Residential and Commercial Settings*, Proceedings of International Association of Science and Technology for Development Conference (IASTED): Power and Energy Systems; Domijan, A., Ed.; ACTA Press: Calgary, AB, 2001; ISBN: 0-88986-317-2.
4. Power Quality Troubleshooting at the Service Panel: Application Note, Fluke Corporation, www.fluke.com/library, retrieved on December 12, 2003.
5. Hucal, M.C. Illuminating energy savings with efficient lighting. In *Environment Design + Construction*; Business News Publishing Company: Troy, MI, 2003; [retrieved on 11/4/2003 from www.edcmag.com].
6. OS-650 Series Energy Conservation and Plant Maintenance Kits: Operator's Manual. Omega Engineering, Inc., 1995. Available from: www.omega.com
7. ThermoView™ Ti30 Manual. Fluke Corporation (Formally Raytek), 2004. Available from: http://us.fluke.com/usen/products/Fluke_Ti30.htm?trck=ti30
8. Bacharach Inc. Why Test?, 1999; retrieved on February 21, 2004 from www.bacharach_inc.com
9. Superior Signal Company, Inc. retrieved on February 21, 2004 from www.SuperiorSignal.com
10. DI-154RS and DI-194RS Starter Kits, DATAQ Instruments, Inc., 2003; retrieved on February 22, 2004 from www.dataq.com

Comp–Day

Daylighting

William Ross McCluney
Florida Solar Energy Center, University of Central Florida, Cocoa, Florida, U.S.A.

Abstract
Daylight illumination of building interiors is an ancient art now benefitting from relatively recent engineering advances. The benefits are numerous and include energy savings and enhanced visual and thermal comfort. The design must avoid overheating and the discomfort and reduced productivity resulting from glare. Good daylighting design can displace electric lighting and reduce air pollution, global warming, and dependence on dwindling supplies of fossil-fuel energy.

INTRODUCTION

Daylighting has been a primary means of interior illumination since the first buildings were constructed. Full use of natural daylight can displace daytime electric lighting energy consumption in buildings, resulting in considerable avoidance of building energy operating costs. Designing for daylight illumination is an ancient art, enhanced considerably in the modern era by improved materials, construction techniques, and computerized design and performance evaluation tools. Artists have known for centuries that natural daylight offers the best illumination for good color rendering of paintings and other objects. Now it is possible in many climates to use larger window areas than in the recent past—for the aesthetic, view, illumination, and health benefits these possess—without the adverse energy consequences large windows once represented.

Strong direct illumination from the solar disk, however, presents an important challenge to the daylighting designer. The sun's motion through the sky means that orientation and shading strategies for minimizing direct beam glare must incorporate the known paths of the sun through the sky each day. Fortunately, several design strategies are available to ease the design process. The hourly energy and illumination performances of windows and other daylighting systems now can be assessed quickly and with modest precision.

ILLUMINATION BASICS

Radiometry and Photometry

Radiometry is a system of language, mathematical formulations, and instrumental methodologies used to describe and measure the propagation of radiation through space and materials. Photometry is a subset of radiometry dealing with radiation in the visible portion of the spectrum. Only radiation within the visible portion of the spectrum, ranging from approximately 380 nm to approximately 720 nm, should be called light. Photometric quantities are defined in such a way that they incorporate the variations in spectral sensitivity of the human eye over the visible spectrum—as a spectral weighting function built into their definition. Though daylight illumination of building interiors deals primarily with photometric quantities, radiometric ones are important in assessing the energy-performance features of daylighting systems.

Photometric quantities may be derived from their spectral radiometric counterparts using the equation below. Let Q_λ be any of the four spectral radiometric quantities (flux, irradiance, intensity, or radiance), and let $V(\lambda)$ be the human photopic spectral luminous efficiency function. The photometric equivalent, Q_v, subscripted with the letter v (for visual) of the radiometric quantity Q_e, subscripted with the letter e (for energy) is the weighted integral of the spectral radiometric quantity over the visible portion of the spectrum.[1]

$$Q_v = 683 \int_{380\ \mathrm{nm}}^{760\ \mathrm{nm}} V(\lambda) Q_\lambda(\lambda) \mathrm{d}\lambda$$

The Paths of the Sun Through the Sky

The sun moves in a predictable way through the sky each day. In the Northern Hemisphere at middle latitudes in the winter, it rises south of due east (azimuth 90) and sets south of due west (azimuth 270). In summer, it rises north of due east and sets north of due west. A plot of solar position vs time on a chart of solar coordinates is called a sunpath chart. Knowledge of solar movement is important in designing daylighting systems. Such systems perform best when they minimize glare and overheating from solar radiant heat gain. The World Wide Web offers several tools for determining the position of the sun in the sky.

Keywords: Daylight; Glazing; Aperture; Illumination; View; Windows; Energy.

A sunpath-chart drawing program is available for free download from the Web site of the Florida Solar Energy Center (www.fsec.ucf.edu).

Proper orientation of buildings and spaces with glazed apertures relative to solar movement is very important in building design for daylighting. Shading devices—including overhangs, side fins, awnings, window reveals, and a variety of exterior and interior shades and shutters—are important tools for the daylighting designer. In addition, new materials and design strategies permit the use of concentrated and piped daylighting systems, using the strong flux from direct beam sunlight to minimize aperture areas while delivering sunlight without glare to spaces somewhat remote from the building envelope.

The Solar Spectrum

Spectra for direct beam and diffuse sky radiation on a vertical wall, with the sun 60° above the horizon on a clear day, are plotted in Fig. 1, along with a scaled plot of the human photopic visibility function. The sky is blue as a result of spectrally selective scattering of light from air molecules. This is seen in the shift of the diffuse sky spectrum in Fig. 1 toward shorter (bluer) wavelengths. Light through windows from blue sky alone does not appear strongly colored to the human observer. One reason is the adaptability of the human visual system. Another reason is the presence of a wide range of colors in sun and sky light, except for the case of sunlight at sunrise and sunset (which is reddish by virtue of the greater mass of intervening atmosphere, which removes some of the blue light in the spectrum). The solar spectrum under an overcast sky has a shape approximately the same as the sum of the direct and diffuse spectra, because scattering of light from water droplets in the atmosphere—the principal mechanism producing light from clouds and overcast skies—has weak spectral selectivity.

Electric lighting is approximately constant in output, in the absence of dimming systems. Daylight is changing constantly. The variability in sun and sky light is both a problem and a benefit in daylighting design. The benefit stems from the positive responses most people feel when experiencing the natural changes in daylight illumination in the absence of glare. The problem comes from the need to design the daylighting system to respond well to the changes. The design process begins with an understanding of the daylight availability for the building site, including the blocking effects of trees, buildings, and other nearby objects. Daylighting systems intended for an area experiencing predominantly overcast skies will be different from those designed for mostly clear-sky conditions. The National Renewable Energy Laboratory in Golden, Colorado in the United States and other national laboratories around the world, including weather bureaus and other such services, can be consulted for information about daylight availability for sites in their jurisdictions. In 1984 the Illuminating Engineering Society of North America published a guide to daylight availability.[2]

Glare

There are two kinds of glare, illustrated in Figs. 2 and 3. With disability glare, light reflecting from the surface of a visual task masks the contrast in that task and degrades the ability to see it. Examples include light reflected from a

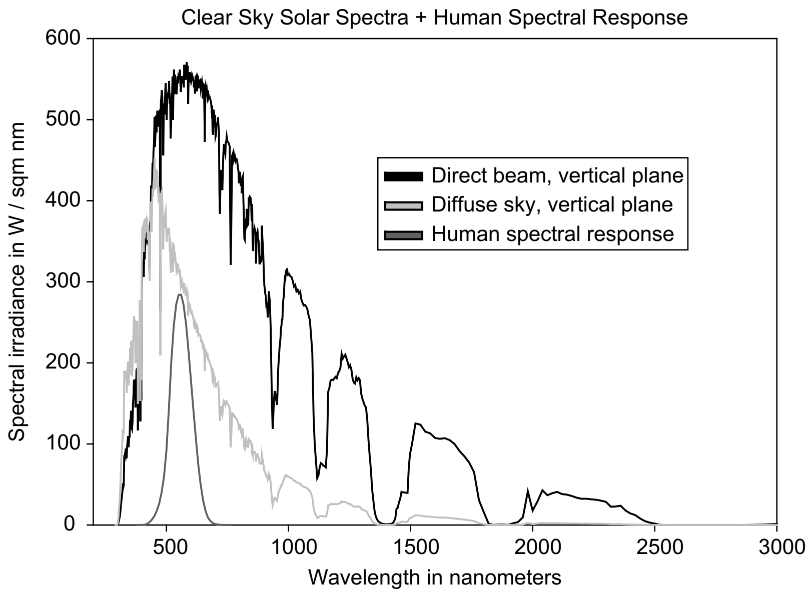


Fig. 1 Direct and diffuse solar spectra on a vertical plane. The sky is clear, and the solar-altitude angle is 60°. Also shown is the photopic spectral luminous efficiency function, whose peak value is 1.0, scaled up for clarity. (Credit: Author.)

Comp–Day

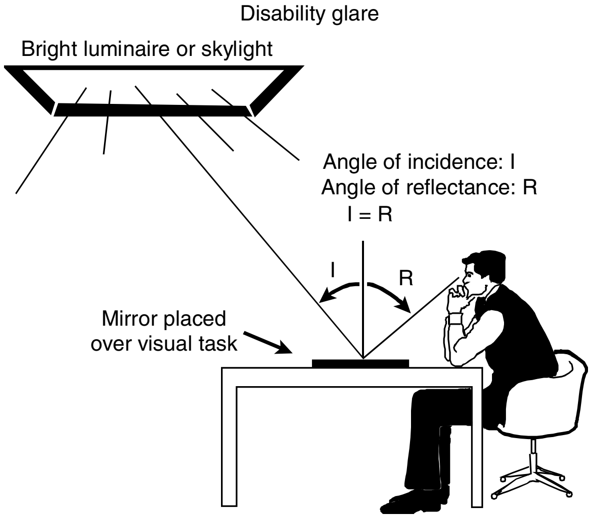


Fig. 2 Disability glare is produced with a strong source of light that masks a visual task, usually by reflection, reducing its contrast and disabling the person's ability to see that task well. Reorienting the task and the glare source often can ameliorate the problem, as can reducing the brightness of the source. (Credit: Florida Solar Energy Center.)

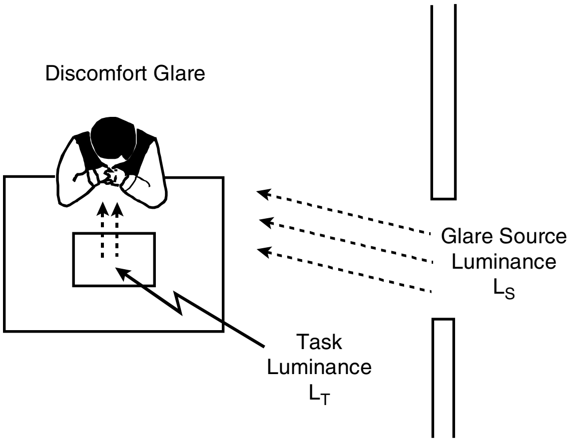


Fig. 3 Discomfort glare is produced by strong light from the side entering the eye without masking the target directly. (Credit: Florida Solar Energy Center.)

computer screen or television set and light reflected from a glossy magazine page. The reflected light reduces the contrast of the image. With a glossy magazine, for example, the reflected glare light can be as strong from the black ink as from the white paper, washing out the text and making it difficult or impossible to read.

Discomfort glare usually results when light entering the eye from the side is much brighter than that coming from the visual task. This extra-bright light is mentally and physically confusing, and can result in visual fatigue, discomfort, and even headaches.

Buildings have several potential sources of glare. Beam sunlight entering the eye directly perhaps is the worst, due to its extreme brightness. Both specularly and diffusely reflected beam sunlight also can produce both kinds of glare. A bright window surrounded by dark walls and furnishings nearly always produces discomfort glare unless the light from the window normally does not enter the eyes, due to the orientation of the visual task. Bare electric lamps, either incandescent or fluorescent, can produce glare, as can poorly designed electric luminaires. A successful lighting design will reduce or minimize the system's potential contribution to both kinds of glare, whether it be a daylighting or an electric lighting system.

A variety of metrics have been devised in attempting to quantify visual comfort in the presence of discomfort-glare sources. Most incorporate terms for the angles of the brighter source from the direct line of sight into the eye.[3] A rule of thumb for reducing discomfort glare is to keep the brightest light source in the visual field from being stronger than a few times the general surround luminance.

Fig. 4 illustrates two contrasting daylighting designs: one that promotes discomfort glare and one that ameliorates it. If the window in both cases looks out on the same uniform sky, it will have the same luminance in both cases. The larger window, therefore, will admit greater overall flux into the room. With moderately high room surface reflectances, this means that the room generally will be brighter and the room surface luminances will be closer to that of the window, producing less tendency toward glare. The same effect can be achieved by reducing the visible transmittance of the window while increasing the electric lighting in the room, reducing the contrast in room brightnesses, but this "solution" calls for more purchased energy.

A final caveat is offered regarding glare. One person's "killer glare" is another person's "sparkle." People living in a region with persistent clouds and long, dark winters may have a much higher tolerance for direct beam admission than those living in a hot region with

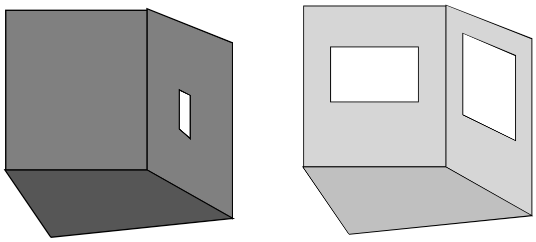


Fig. 4 Comparison of two daylighting designs. The one on the left has a relatively small window in a room with low surface reflectances, resulting in inadequate illumination, a gloomy appearance, and a propensity for discomfort glare. The one on the right, having larger window areas and higher surface reflectances, is brighter and less prone to glare. (Credit: Author.)

predominantly clear skies and high direct beam solar flux levels. It is important to use direct sunlight entry judiciously and with care to prevent the adverse impacts that can result. Some daylighting systems rely almost exclusively on direct beam sunlight, but these are designed to distribute the concentrated beams widely and diffusely, with minimal glare impact.

DAYLIGHTING DESIGN

The goals of good daylighting include the provision of good-quantity and good-quality daytime interior illumination and view, coupled with high visual comfort for occupants. Happy people are productive people. It is necessary to prevent common problems such as unwanted glare and overheating from excessive direct beam illumination. In the process of providing good-quality illumination of adequate quantity, it is desirable to design the system so as to displace as much daytime electric lighting as possible, to minimize the energy costs of building operation.

Some traditional means of admitting daylight into building spaces are drawn schematically in Fig. 5.

Additional approaches using light pipes and other beam-manipulation strategies can also be utilized. Several of these approaches are illustrated schematically in Fig. 6.

BUILDING OCCUPANCY

The most efficient buildings are unoccupied ones, with all the building's energy services turned off. Consequently, the more a building is occupied and using energy, the greater are the opportunities for energy savings. Daylighting design saves energy through averted electrical energy costs. If sufficient daylighting already is available in a building, however, additional daylighting could increase energy costs. On the other hand, health and comfort benefits often are sufficient to justify the introduction of some additional daylight if the extra energy costs are modest, limited by the use of high-performance windows.

If people are seldom in a building during the daylight hours, the need for illumination is minimal, and daylighting can't save much energy. This typically is the case for most residences. Exceptions include residences occupied by retirees and others not working outside the home in the daylight hours, and those daytime-occupied homes suffering inadequate daylighting. In these cases, added daylighting makes good energy sense, and offers additional visual comfort and psychological benefits. For offices and other buildings fully occupied during daylight hours, daylight illumination often saves more building energy than any other single strategy, and the attendant increases in worker productivity add further to the benefits.

DAYLIGHTING SIMULATION AND MODELING

A flow chart showing the connections among various aspects of daylighting system energy and illumination performance is shown in Fig. 7.

Over the past two centuries, a variety of methodologies has been developed for predicting both the energy and the illumination performances of daylighting systems. With the advent of the fast personal computer, most of these methods have given way to sophisticated new computer

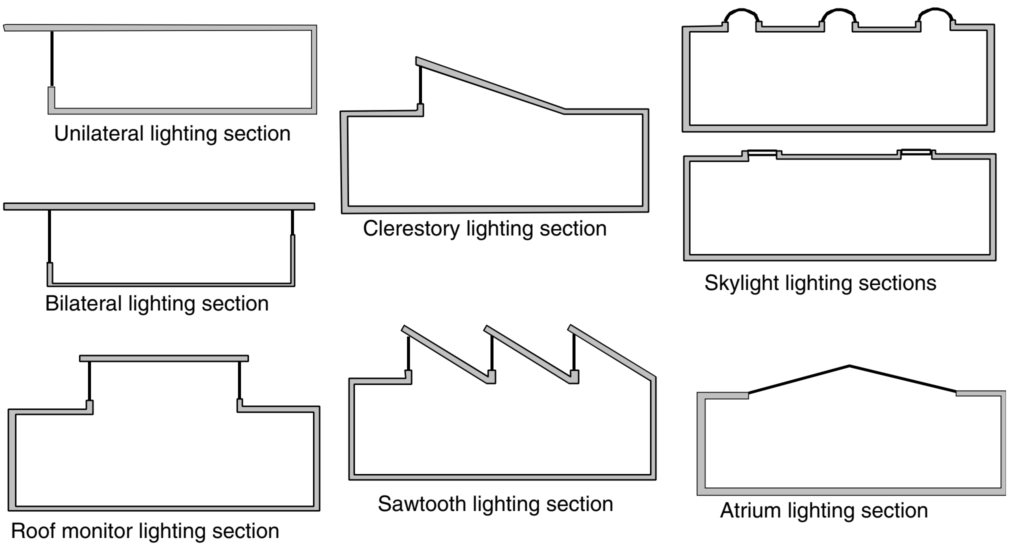


Fig. 5 Section views of several ways daylight can be admitted into buildings. (Credit: *IESNA Lighting Handbook*. 9th ed. Illuminating Engineering Society of North America, 2000.)

Comp–Day

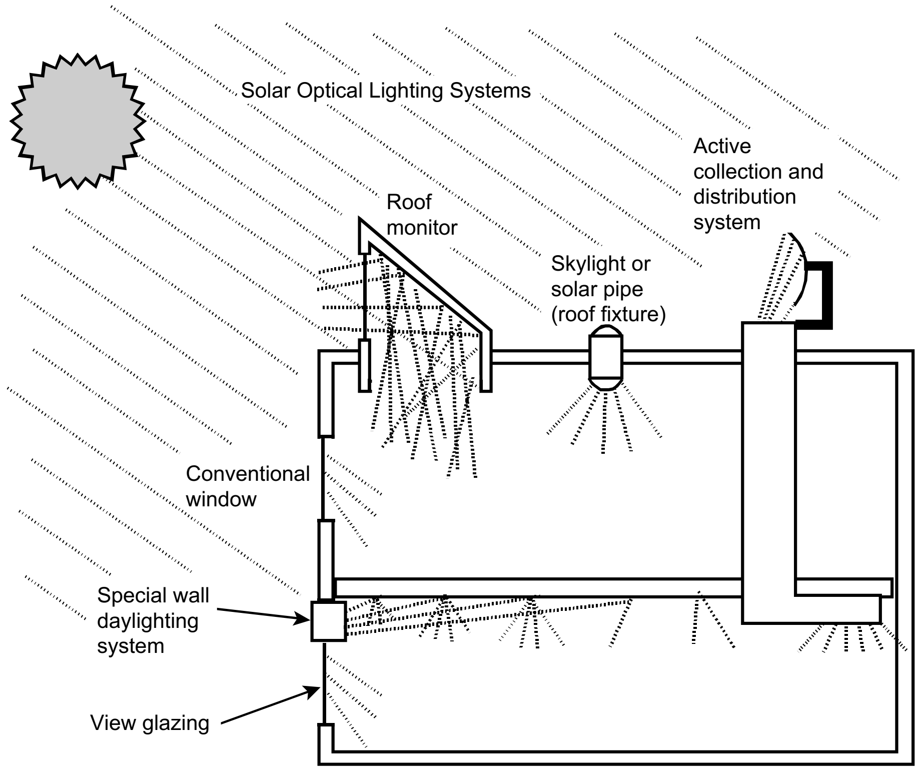


Fig. 6 Illustration of the variety of ways daylight can be admitted into building spaces for controlled illumination of the interior. (Credit: Author.)

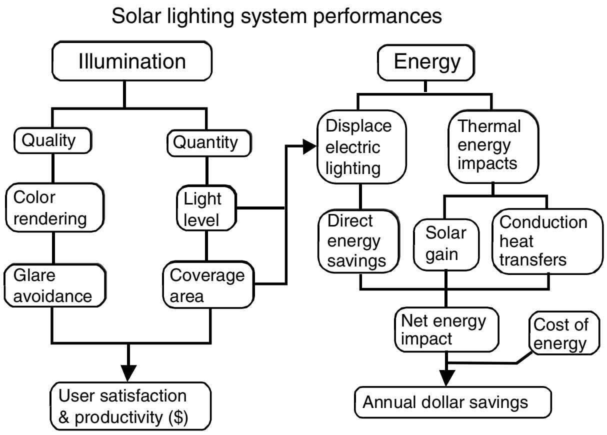


Fig. 7 Connection pathways for the energy and illumination performance components of daylighting systems. (Credit: Author.)

tools for determining the performance of both electric and daylighting systems. The new tools include commercially available computer programs and public-domain software. A Web search on such terms as daylighting design software and architectural lighting design should provide links to many sites describing these computer programs.

Another important tool is scale-model testing. Because illumination scales upward and downward well, small-scale models of buildings can be constructed of inexpensive materials. Light sensors can be placed inside these models to measure light levels when the models are placed in either simulated or natural daylight. Cameras also can be employed to image the light distributions inside the building's daylit rooms, under varied sun and sky conditions, to assess performance visually.

CONCLUSION

Humans need connections with the outdoors. This need is built into our genetic makeup; it promotes health and a sense of well being, and makes us happier and more productive. Even photographs of nature on the wall have proved to be helpful.

According to Judith Heerwagen, principal of J.H. Heerwagen and Associates, and senior scientist at the Pacific Northwest National Laboratory in Seattle, in our evolutionary past, information about our environment had a pronounced influence on survival and health. Changes in daylight provided time cues and assessment of cloud formations for information about future weather conditions. These events influenced our ancestors' daily decisions, such as where to sleep at night and where to look for food next week. Loss of illumination from and the view to the outdoors have been implicated in the poor recovery of patients in windowless intensive care units.

"Once you start thinking about it, [daylighting] design makes perfect sense," Heerwagen has written. "We didn't evolve in a sea of gray cubicles."[4]

Daylighting offers a number of benefits to building owners and occupants. Cool, natural daylight has good color rendering; it is healthy and offers clear psychological benefits. Daylighting can displace electric lighting, saving energy and reducing air pollution, global warming, and our dependence on dwindling supplies of fossil-fuel energy.

Glossary

Brightness: Brightness is a subjective term with no universally acceptable quantitative definition. It refers to a perception of the strength of illumination received by the human eye. It may be used in general characterizations of the appearances of different sources and objects. Luminance approximately characterizes the strength of illumination emanating from an object or source, as perceived by the eye.

Illuminance, E_v: The area density of luminous flux; the luminous flux per unit area at a specified point in a specified surface that is incident on, passing through, or emerging from that point in the surface (units: $\text{lm m}^{-2} = \text{lux}$).

Irradiance, E_e: The area density of radiant flux; the radiant flux per unit area at a specified point in a specified surface that is incident on, passing through, or emerging from that point in the surface (units: W m^{-2}).

Luminance, L_v: The area and solid angle density of luminous flux; the luminous flux per unit projected area and per unit solid angle incident on, passing through, or emerging from a specified point in a specified surface, and in a specified direction in space (units: $\text{lm m}^{-2}\ \text{sr}^{-1} = \text{cd m}^{-2}$).

Luminous flux, Φ_v: The $V(\lambda)$-weighted integral of the spectral flux Φ_λ over the visible spectrum (units: lumen or lm).

Luminous intensity, I_v: The solid angle density of luminous flux; the luminous flux per unit solid angle incident on, passing through, or emerging from a point in space and propagating in a specified direction (units: $\text{lm sr}^{-1} = \text{cd}$).

Photopic spectral luminous efficiency function, $V(\lambda)$: The standardized relative spectral response of a human observer under photopic (cone vision) conditions over the wavelength range of visible radiation.

Radiance, L_e: The area and solid angle density of radiant flux; the radiant flux per unit projected area and per unit solid angle incident on, passing through, or emerging from a specified point in a specified surface, and in a specified direction in space (units: $\text{W m}^{-2}\ \text{sr}^{-1}$).

Radiant flux, Φ_e: The time rate of flow of radiant energy (units: watt).

Radiant intensity, I_e: The solid angle density of radiant flux; the radiant flux per unit solid angle incident on, passing through, or emerging from a point in space and propagating in a specified direction (units: W sr^{-1}).

Radiation luminous efficacy, K_r: The ratio of luminous flux in lumens to radiant flux (total radiation) in watts in a beam of radiation (units: lumen/watt).

Spectral radiometric quantities: The spectral "concentration" of quantity Q, denoted Q_λ, is the derivative $dQ/d\lambda$ of the quantity with respect to wavelength λ, where Q is any one of: radiant flux, irradiance, radiant intensity, or radiance (units: same as that of quantity Q per nm).

System luminous efficacy, K_s: The ratio of luminous flux in lumens delivered in a space to the electrical consumption of the lighting system delivering that flux (units: lumen/watt).

Visible transmittance, T_v or VT: The ratio of transmitted to incident illuminance on a glazing system; a unitless quantity.

Comp–Day

REFERENCES

1. McCluney, R. *Introduction to Radiometry and Photometry*, Artech House: Boston, MA, 1994.
2. IES Calculation Procedures Committee Recommended practice for the calculation of daylight availability. J. IES **1984**, July, 3-11–3-13.
3. Kaufman, J.E.; Christensen, J.F., Eds. In *IES Lighting Handbook, 1984 Reference Volume*, Illuminating Engineering Society of North America: New York, 1984.
4. Heerwagen, J.H. The Psychological Aspects of Windows and Window Design. EDRA 21 Symposium, Environmental Design Research Association, 1990.

Demand Response: Commercial Building Strategies☆

David S. Watson

Sila Kiliccote

Naoya Motegi

Mary Ann Piette

Commercial Building Systems Group, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.

Abstract

This paper describes strategies that can be used in commercial buildings to temporarily reduce electric load in response to electric grid emergencies in which supplies are limited or in response to high prices that would be incurred if these strategies were not employed. The DR strategies discussed herein are based on the results of three years of automated DR field tests in which 28 commercial facilities with an occupied area totaling over 11 million ft^2 were tested. Although the DR events in the field tests were initiated remotely and performed automatically, the strategies used could also be initiated by on-site building operators and performed manually, if desired. While energy efficiency measures can be used during normal building operations, DR measures are transient; they are employed to produce a temporary reduction in demand. Demand response strategies achieve reductions in electric demand by temporarily reducing the level of service in facilities. Heating, ventilating and air conditioning (HVAC), and lighting are the systems most commonly adjusted for DR in commercial buildings. The goal of DR strategies is, to meet the electric shed savings targets while minimizing any negative impacts on the occupants of the buildings, or the processes that they perform. Occupant complaints were minimal in the field tests. In some cases, "reductions" in service level actually improved occupant comfort or productivity. In other cases, permanent improvements in efficiency were discovered through the planning and implementation of "temporary" DR strategies. The DR strategies that are available to a given facility are based on factors such as the type of HVAC, lighting, and energy management and control systems (EMCS) installed at the site.

BACKGROUND

Power requirements on the electric grid are in constant flux, based on the demand of the devices connected to it. This demand varies based on time of day, weather, and many other factors. Traditionally, the supply is varied to meet the demand by increasing or decreasing electric generation capacity. Conversely, demand response (DR) can be defined as short-term modifications in customer end-use electric loads in response to dynamic price and reliability information.

As electric demand increases, generation costs increase in a non-linear fashion. A price spike caused by high demand on a hot summer afternoon would be an example of price information that might be used to initiate short-term modifications in customer end-use electric loads. A scenario in which a power plant failed unexpectedly would be an example of where short-term modifications in customer end-use electric loads could help other on-line plants manage the demand thereby increasing system reliability and avoiding blackouts.

Many electric utilities across the United States have implemented programs that offer financial incentives to ratepayers who agree to make their electric loads more responsive to pricing and/or reliability information. These programs are most prevalent for commercial and industrial customers in utility districts with known capacity or transmission constraints.

Recent studies have shown that customers have limited knowledge of how to develop and implement DR control strategies in their facilities.[2] Another barrier to participation in DR programs is the lack of systems that help automate the short-term modifications or strategies required during DR events.

This paper focuses on strategies that can be used to enable DR in commercial buildings (i.e., to make short-term modifications to their end-use equipment).

RESULTS OF FIELD TESTS

The strategies discussed herein are based on the results of a series of field tests conducted by the PIER Demand

☆ This paper was originally presented as "Strategies for Demand Response in Commercial Buildings" at the 2006 ACEEE Summer Study on Energy Efficiency in Buildings Conference, August 13–18, 2006 in Pacific Grove, California. Reprinted with Permission. ACEEE is the American Council for an Energy Efficient Economy, located in Washington, DC.

Keywords: Demand response; Electric load reduction; Load control; Peak load control; Peak load reduction; HVAC load reduction; Lighting load reduction; Global temperature adjustment.

Table 1 Average and maximum peak electric demand savings during automated demand response (DR) tests

Results by year	Number of sites	Duration of shed (h)	Average savings (%)	Maximum savings (%)
2003	5	3	8	28
2004	18	3	7	56
2005	12	6	9	38

Response Research Center. While the tests focused on fully automated electric DR, some manual and semi-automated DR was also observed. The field tests included 28 facilities, 22 of which were in Pacific Gas and Electric territory. The other sites were located in territories served by Sacramento Municipal Utility District, Southern California Edison, City of Palo Alto Utilities and Wisconsin Public Service. The average demand reductions were about 8% for DR events ranging from 3 to 6 h.

Table 1 shows the number of sites that participated in the 2003, 2004, and 2005 field tests along with the average and maximum peak demand savings. The electricity savings data are based on weather sensitive baseline models that predict how much electricity each site would have used without the DR strategies. Further details about these sites and the automated DR research are available in previous reports.[4,5]

Fig. 1 shows the various DR strategies that were used in field tests and the frequency of each. The tests included building types such as office buildings, a high school, a museum, laboratories, a cafeteria, data centers, a postal facility, a library, retail chains, and a supermarket. The buildings range from large campuses, to small research and laboratory facilities.

Fig. 2 shows the various DR strategies that were used in field tests and the Demand Saving Intensity (W/ft^2) by Shed Strategy. The values shown are average savings over 1 h. Though the sample size is not large enough to generalize shed savings by strategy, it is clear that each of the three shed categories listed has the potential to shed about 0.5 W/ft^2. Most of the DR heating, ventilating and air conditioning (HVAC) strategies we have examined provide considerably greater savings on hotter days and the data in Fig. 2 were from a mild day. Lighting strategies are not weather dependent.

CONCEPTS AND TERMINOLOGY

Energy Efficiency

Energy efficiency can lower energy use without reducing the level of service. Energy efficiency measures are part of normal operations to permanently reduce usage during peak and off-peak periods. In buildings, energy efficiency is typically achieved through efficient building designs, the use of energy efficient equipment, and through efficient building operations. Since energy efficiency measures are a permanent part of normal operations, they are typically considered separate from DR which involves short term modifications to normal operations. However, some energy efficiency measures such as the use of variable frequency drives (VFDs) on electric motors can enable both energy efficiency and temporary DR modes when called to do so.

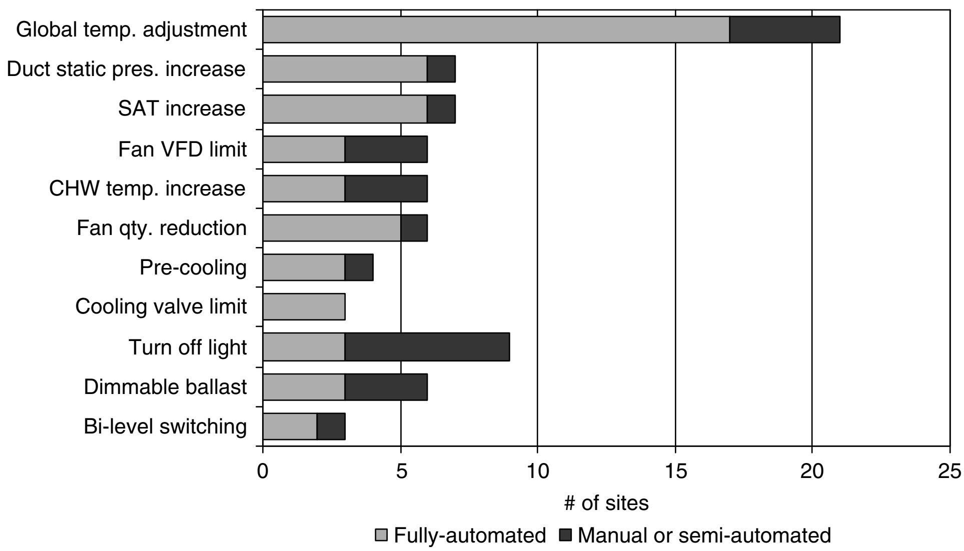


Fig. 1 Frequency of various demand response (DR) strategy usage.

Dem–Dry

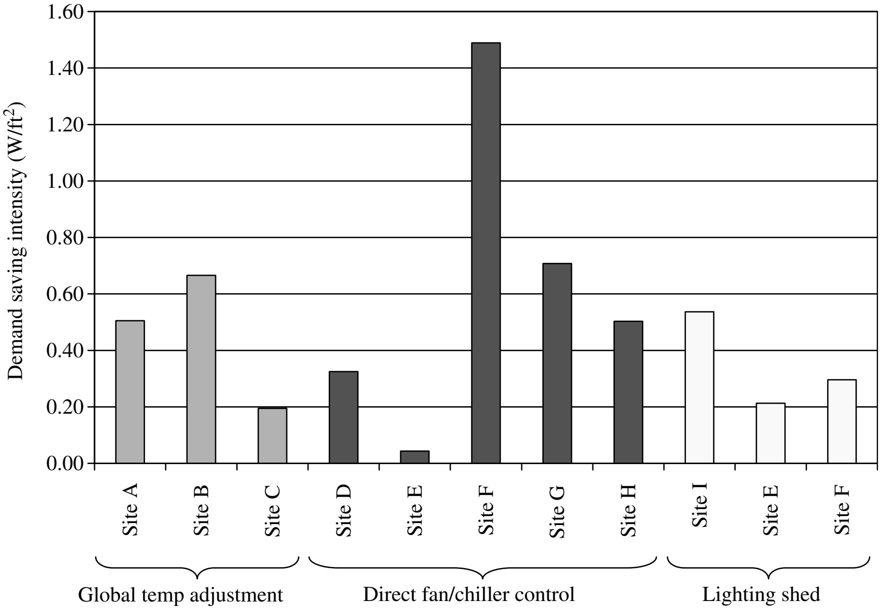


Fig. 2 Demand saving intensity (W/ft^2) by shed strategy on November 5, 2004.

Daily Peak Load Management

Daily peak load management is done in many buildings to minimize peak demand charges and time-of-use rates. Strategies that temporarily modify the operation of HVAC or lighting systems are often used to implement daily peak load management. Decisions about when to initiate daily peak load management are typically made by on-site staff or on-site automated equipment.

Demand Shifting

Demand shifting is achieved by changing the time that electricity is used. Thermal energy storage is an example of a demand shifting technology. Thermal storage can be achieved with active systems such as chilled water or ice storage, or with passive systems such as pre-cooling the mass of a building.[6] Both daily peak load management and demand shifting are typically done to minimize peak demand and time-of-use rate charges.

Demand Response

Demand response can be defined as short-term modifications in customer end-use electric loads in response to dynamic price and reliability information. Demand response events are dynamic and temporary. They are driven by factors such as low electricity reserves, warm weather and grid conditions.

One of the key components of DR is that the pricing and reliability information known at the grid system or utility level must be transmitted and translated into load reducing actions at the end-use sites. Signaling methods used to inform facility operators of upcoming DR events include: phone calls, pagers, text messages, and e-mail messages. Control signals are also used in some systems for direct signaling to energy management and control systems (EMCS) and control of electric loads. These digital control signals are broadcast using radio transmissions, power-line communications and the Internet.

Demand response can be implemented using various levels of automation. *Manual Demand Response* is performed by facilities staff physically turning off electric equipment after receiving notification of an upcoming DR event. *Semi-Automated Demand Response* is similar, but reduces facilities staff labor through use of a centralized control system with pre-programmed DR strategies. *Fully-Automated Demand Response* enables remotely generated event initiation signals to control loads directly or to initiate pre-programmed DR strategies at the site. Though Fully-Automated DR is capable of functioning without human intervention, it is recommended that facility operators are kept informed of the process and have the ability to “opt-out” of a DR event, if desired.

Reduction in Service

Demand response strategies achieve reductions in electric demand by temporarily reducing the level of service in facilities. Heating, ventilating and air conditioning, and lighting are the systems most commonly adjusted to achieve DR savings in commercial buildings. The goal of DR strategies is to meet the electric shed savings targets while minimizing any negative impacts on the occupants of the buildings or the processes that they perform. Occupant complaints were minimal in the field tests. In some cases, “reductions” in service level actually improved occupant comfort or productivity. Such cases can be caused by over-cooling that occurs in some buildings during normal operation. In other cases, permanent improvements in efficiency were discovered through the planning and implementation of "temporary" DR strategies. The DR strategies that are available to a given facility are based on factors such as the type of HVAC, lighting, and EMCS installed at the site.

Shared Burden

Demand response strategies that share the burden evenly throughout the facility are least likely to have negative effects on building occupants. For example, if it were possible to reduce lighting levels evenly throughout an entire facility by 25% during a DR event, impact to occupants may be minimal. However, turning off all of the lights in one quadrant of an occupied space would not be acceptable. In HVAC systems, strategies that reduce load evenly throughout all zones of a facility are superior to those that allow certain areas (such as those with high solar gains) to substantially deviate from normal temperature ranges.

By combining savings from sheds in HVAC and lighting (and other loads, if available), the impact on each system is minimized and the savings potential is increased.

Closed Loop Control

Comfort is maintained in modern buildings through the use of closed loop control of HVAC systems. Sensors are used to measure important parameters such as temperature and pressure. Controllers adjust actuators such as dampers or valves to maintain the desired setpoints for those parameters. The effect of the actuators on the controlled zone or system is measured by the sensor, hence "closing the control loop." Control sub-systems for which there is no feedback from sensors are known as "open loop" controls.

In order to maintain predictable and managed reductions of service during DR events, strategies should maintain the use of closed loop controls in HVAC systems.

Granularity of Control

For the purpose of DR control in buildings, the concept of granularity refers to how much floor area is covered by each controlled parameter (e.g., temperature). In HVAC systems, the ability to easily adjust the temperature setpoint of each occupied space is a highly granular way to distribute the DR shed burden throughout the facility. Less granular strategies such as making adjustments to chillers and other central HVAC equipment can provide effective shed savings, but can cause temperature in some zones to drift out of control. Granularity of control can also allow building operators to create DR shed behaviors that are customized for their facility. An example of this would be to slightly increase all office zone temperature setpoints, but leave computer server room setpoints unchanged.

Resolution of Control

In HVAC systems, parameters are controlled with great resolution. In many systems temperature setpoints can be adjusted by as little as 0.1°F. Although some modern lighting ballasts can adjust individual lamps in less than 1% increments, most commercial lights are only capable of being turned on or off. Additional information is provided in the "Lighting Based DR Strategies" section below.

Rebound

At the end of each DR event, the effected systems must return to normal operation. When lighting strategies are used for DR, normal operation is regained by simply re-enabling all lighting systems to their normal operation. Lights will come back on as commanded by time clocks, occupancy sensors or manual switches. There is no reason for lighting power to jump to levels that are higher than normal for that period.

However, without special planning HVAC systems tend to use extra energy following DR events in order to bring systems back to normal conditions. Extra energy is used to remove heat that is typically gained during the reduced service levels of the DR event. This post DR event spike in demand is known as "rebound." To minimize high demand charges and to reduce negative effects to the electric grid, rebound should be reduced or minimized through use of a strategy that provides a graceful return to normal operation. The simplest case is where the DR event ends or can be postponed until the building is unoccupied. If this is not possible, strategies that allow HVAC equipment to slowly ramp up or otherwise limit power usage during the return to normal period should be used.

HVAC BASED DR STRATEGIES

Heating, ventilating and air conditioning systems can be an excellent resource for DR shed savings for several reasons: (1) HVAC systems create a substantial electric load in commercial buildings, often more than one-third of the total; (2) the "thermal flywheel" effect of indoor environments allows HVAC systems to be temporarily unloaded without immediate impact to the building occupants; and (3) it is common for HVAC systems to be at least partially automated with EMCSs.

However, there are technical challenges to using commercial HVAC systems to provide DR sheds. These systems are designed to provide ventilation and thermal comfort to the occupied spaces. Operational modes that provide reduced levels of service or comfort are rarely included in the original design of these facilities. To provide reliable, repeatable DR sheds it is best to pre-plan and automate operational modes that will provide DR savings. The use of automation will reduce labor required to implement DR operational modes when they are called. In addition, timeliness of the response will typically be improved.

Heating, ventilating and air conditioning based DR strategies recommended for a given facility, vary based on the type and condition of the building, mechanical equipment and EMCS. Based on these factors, the best DR strategies are those that achieve the aforementioned goals of meeting electric shed savings targets while minimizing negative impacts on the occupants of the buildings or the processes that they perform. The following DR strategies are prioritized so as to achieve these goals:

1. Global temperature adjustment (GTA) of zones
2. Centralized adjustments to the air distribution and/or cooling systems.

All HVAC based DR strategies outlined in this paper allowed zone temperatures to drift outside of normal ranges. However, the rate at which the temperatures drifted was well below the rate of Acceptable Temperature Change defined in ASHRAE Standard 55-2004. Demand response strategies used to return the HVAC system to normal operation should be designed for a similarly gradual rate of change. In addition to the comfort benefits outlined in the ASHRAE standard, strategies that slowly return the system to normal have the additional benefit of limiting rebound spikes as described previously.

GLOBAL TEMPERATURE ADJUSTMENT OF ZONES

Description

Global temperature adjustment of occupied zones is a feature that allows commercial building operators to easily adjust the space temperature setpoints for an entire facility from one command from one location. Typically, this is done from a screen on the human machine interface (HMI) to the EMCS. In field tests, GTA was shown to be the most effective and least objectionable strategy of the five HVAC shed strategies tested.[4] It is most effective because it reduces the load of all associated air handling and cooling equipment. It is least objectionable because it shares the burden of reduced service level evenly between all zones. Global temperature adjustment based DR strategies can be implemented either manually by building operators or automatically based on remote signals.

Typical Implementation

Global temperature adjustment is typically implemented by broadcasting a signal from the central EMCS HMI server to the all final space temperature control devices distributed throughout the facility. Upon receipt of a global signal from the central EMCS server, the final space temperature control devices interprets the signal and reacts accordingly (e.g., DR Mode Stage-1 means increase space cooling setpoints 3°F and decrease space heating setpoints 3°F).

Final space temperature control devices suitable for GTA include:

- Space temperature controllers that adjust variable air volume (VAV) terminal box dampers (all types) (e.g., VAV boxes).
- Space temperature controllers that adjust hot water heating coil valves or chilled water cooling coils (e.g., fan coil units, CAV multi-zone heating and cooling coil valves).
- Space temperature controllers that adjust capacity of heat pumps or direct expansion (DX) units.

To avoid an unwanted increase in heating energy, heating setpoints should remain the same or be reduced during GTA mode.

Mode Transitions

In the most basic implementation, upon receipt of a DR signal the GTA enabled system will increase space cooling setpoints in one or two steps (two step increase shown in Table 2). Upon entering a DR mode (e.g., moderate shed), the global temperature setpoints will be increased and load on the air distribution and cooling systems will decrease.

More advanced implementations can adjust setpoints to follow linear or exponential curves.[6] Though more difficult to program, these strategies can provide added flexibility in creating shed profiles that are customized to provide optimal consistency or duration for a given facility.

Decay of Shed Savings

Over time, internal and external heat gains will increase zone temperatures until they exceed the new DR setpoints, causing fan and cooling systems to ramp back up. This phenomenon, known as "decay" of shed savings, can be prevented by further increasing the zone cooling setpoints to new levels (e.g., high shed). After a certain time duration, which varies by building type, weather and other factors, the shed savings will decay to the point where additional setpoint increases are not viable in an occupied building. In field tests, successful sheds of up to 6 h have been performed without substantial impact on commercial building occupants.

Table 2 Global temperature adjustment (GTA) setpoint adjustment—example of absolute and relative implementations

Demand response (DR) mode	Absolute space temperature cooling setpoints (°F)	Relative space temperature cooling setpoints
Normal	74 (globally)	Varies per zone
Moderate shed	76	Normal +2°F
High shed	78	Normal +4°F

Absolute vs Relative Implementation

Global temperature adjustment may be implemented on either an absolute or relative basis (Table 2). An absolute implementation of GTA allows the operator to set the space temperature setpoints for the entire facility to absolute values (e.g., heating setpoints at all final space temperature control devices = 68°F and cooling setpoints at all final space temperature control devices = 76°F). A relative implementation of GTA allows the operator to adjust the space temperature setpoints for the entire facility to new values that are offset from the current values by a relative amount (e.g., heating setpoints at all final space temperature control devices should decrease 2°F from current values and cooling setpoints should increase 2°F from current values). A relative implementation of GTA is best suited for sites where "normal" setpoints vary throughout the facility. It ensures that temperature will not deviate more than a fixed amount from the customized normal setpoint for each zone.

Factory vs Field Implementations of GTA

Several manufacturers offer GTA as a standard feature in their EMCS products. In field tests, sites that used EMCS products from these vendors provided some of the largest sheds and required the least amount of set-up labor. For sites that have EMCS controlled space temperature zones, but lack GTA, it can typically be added in the field. To add GTA to an existing site, each EMCS zone controller must be programmed to "listen" for global GTA commands from the central EMCS system. In addition, the central system must be programmed to send GTA commands to all relevant zone controllers on the EMCS digital network. Typically GTA commands are sent in a global broadcast to all controllers simultaneously.

Impediments to Using GTA Strategy

In field tests, sites that used HVAC shed strategies other than GTA usually did so because that feature was not available at their site. Reasons that GTA is not available include:

- Space temperature not controlled by EMCS (e.g., use of pneumatic controls in occupant zones).
- Space temperature is controlled by EMCS, but space temperature controllers do not include the GTA feature. (i.e., EMCS can adjust space temperature setpoints in each zone individually, but not globally). Adjusting each zone individually is more time consuming and error prone to use for DR purposes.

Evaluation of Global Temperature Adjustment of Zones

While the GTA DR strategy reduces the service level of the occupied spaces, it does so using a closed-loop control strategy in a highly granular fashion. This causes the DR shed burden to be evenly shared between all building occupants and keeps all zones under control. Since none of the zones are starved for airflow, there is no risk of ventilation rates dropping below specified design levels. If GTA of zones is available, the HVAC DR shed strategy recommended for commercial buildings.

AIR DISTRIBUTION AND COOLING SYSTEM ADJUSTMENT

In systems for which the aforementioned GTA of zones is not an option, strategies that make temporary adjustments to the air distribution and/or mechanical cooling systems can be employed to enable DR. Depending on the mechanical systems in place at a given facility, the following DR strategies may be used.

Duct Static Pressure Setpoint Reduction

For variable air volume systems, duct static pressure (DSP) is typically measured in the supply duct. The EMCS modulates the speed of the fan or the position of inlet guide vanes (IGV) to maintain a defined DSP setpoint at the measured location. The "normal" DSP SP at the measured point should be high enough to provide enough pressure for each terminal VAV box to function properly. In an ideal system, the DSP SP would be set just high enough to meet the pressure requirements of the VAV terminal box of greatest demand. But since the box is of greatest demand, and its associated pressure requirement are in constant flux, sub-optimal, yet substantially simpler strategies are usually used to control DSP. Typically DSP is measured at a single location about two-third of the way down the duct system. The DSP SP is set to a fixed value that is high enough to meet the needs of the box of greatest demand during design load conditions. During less demanding conditions energy is wasted due to losses associated with the DSP SP being higher than necessary to meet the demands of the VAV terminal boxes.

Fan energy and cooling energy can be reduced during DR events by reducing the DSP setpoint. This strategy is effective for three reasons:

1. The "normal" DSP SP is often higher than necessary. By reducing the DSP SP, some shed savings is provided without any reduction in comfort or service to the occupants.
2. Additional shed savings occurs when the DSP SP is set low enough to cause some VAV terminal boxes to "starve" from lack of air pressure. This reduction in service causes less air flow through the fans. There is some risk of ventilation rates dropping below specified design levels in some areas using this strategy.
3. When airflow drops below levels necessary to cool the space, electric load on the cooling system also drops.

Fan Speed Limit

Like DSP setpoint reduction mentioned above, this DR strategy is relevant to fans with VFD. During the DR event, the speed of the VFD is limited to a fixed value. To be effective, the fixed value must be lower than if it were allowed to operate under normal closed loop conditions. Fan speed limiting saves energy for the same reasons as DSP setpoint reduction. Its effect on the air distribution systems and associated occupied zones is somewhat less predictable because of the open-loop nature of the control. Fan speed limits may be useful as part of other DR strategies such as cooling system adjustments described below. This strategy may also be used on fans with IGV.

Fan Quantity Reduction

For constant air volume fan systems, the only way to reduce fan energy is by turning fans off completely. This is obviously a severe reduction in service, although it may be of some use in common areas served by multiple fans. If such a strategy is used, it should be noted that cooling energy in the fans that remain on will increase to make up for those that are off.

Increase Supply Air temperature

This strategy saves mechanical cooling energy. In packaged DX units and heat pumps, the savings will be achieved at each unit. For air handlers with cooling coils, the savings will occur at the central cooling plant. In either case, care must be taken to avoid increased fan energy in VAV systems due to increased air flow. This effect can be prevented by limiting fan speeds to levels in use prior to the increase in supply air temperature.

Central Chiller Plants

Most modern centrifugal, screw and reciprocating chillers have the capability of reducing their demand for power. This can be done by raising the chilled water supply temperature setpoint or by limiting the speed, capacity, the number of stages or current draw of the chiller. The quantity of chillers running can also be reduced in some plants.

Evaluation of Air Distribution and Cooling System Adjustment Strategies

While effective in terms of the ability to achieve load reductions, the use of centralized adjustments to air distribution systems and/or mechanical cooling systems for DR purposes have some fundamental drawbacks. In these strategies, the DR burden is not shared evenly between all the zones. Centralized, changes to the air distribution System and/or mechanical cooling systems allow zones with low demand or those that are closer to the main supply fan to continue to operate normally and hence not contribute toward load reduction in the facility. Zones with high demand, such the sunny side of the building or zones at the ends of long duct runs can become starved for air or otherwise go completely out of control. Centralized HVAC DR shed strategies can allow substantial deviations in temperature, airflow and ventilation rates in some areas of a facility. Increased monitoring of occupied areas should be conducted when using these strategies.

LIGHTING BASED DR STRATEGIES

Lighting systems offer great promise as a resource for DR shed savings for several reasons: (1) Lighting systems create a substantial electric load in commercial buildings, often more than 30% of the total; (2) lighting has no rebound effect during the transition from DR events to normal operations; and (3) the lighting systems in many California commercial buildings already have bi-level switching in place. Usually, this enables one-third or two-third or the lights in a given office to be turn off, leaving sufficient light for egress and many common office tasks.

However, there are major impediments to the use of lighting systems for DR: (1) Few office buildings have centralized control of lighting systems;[3] (2) even buildings with centralized lighting controls are not necessarily zoned in a way that would allow a reduction in lighting service that is adequate for occupancy.

Granularity of control is a very important factor in determining the usefulness of lighting systems for DR. The following lists five types of lighting systems from most coarse to most fine granularity: Zone Switching, Fixture Switching, Lamp Switching, Stepped Dimming, Continuous Dimming.

Zone Switching

In areas that are unoccupied or are illuminated by windows or other sources, entire lighting zones can be switched off for DR purposes. In some cases, this strategy can be applied to common spaces such as lobbies, corridors, and cafeterias.

Fixture/Lamp Switching

Fixture or lamp switching can be done by bi-level switching. California's Title 24 Energy Efficiency Building Standard, requires multiple lighting level controls in all individual offices built since 1983. With bi-level switching, each office occupant is provided with two wall switches near the doorway to control their lights. In a typical installation, one switch would control one-third of the fluorescent lamps in the ceiling lighting system, while the other switch would control the remaining two-third of the lamps. This allows four possible light levels: OFF, one-third, two-third and FULL lighting. The 2001 standards state that bi-level switching can be achieved in a variety ways such as:

- Switching the middle lamps of three lamp fixtures independently of outer lamps (lamp switching)
- Separately switching "on" alternative rows of fixtures (fixture switching)
- Separately switching "on" every other fixture in each row (fixture switching)
- Separately switching lamps in each fixture (lamp switching).

Step Dimming

Through the use of ON/OFF switches, controls to regulate the level of electrical light, step dimming is a popular energy-saving retrofit solution for applications where existing fixtures are not equipped with dimming ballasts. Stepped dimming is often called bi-level dimming because the strategy often involves two levels of light output, usually 100 and 50%. However, if more flexibility is required, stepped dimming can involve three levels of light output.

Continuous Dimming

Continuous dimming ballasts allow light output to be gradually dimmed over the full range, from 100 to 10% (fluorescent) or 100 to 50% (HID). These lighting systems provide an excellent resource for DR purposes. These systems allow the lighting load to be reduced so gradually that modest changes may not even be noticed by building occupants.[1] Since the amount of reduction is continuously variable, specific DR shed goals can be achieved using straightforward strategies. As with GTA, shed strategies using continuously dimming lighting can be implemented in an absolute (building-wide) or relative fashion.

In addition to their use for DR, dimmable ballasts can be used in the design of energy efficient systems that reduce electric light requirements when daylight is available. Also, when dimming is available, for many tasks occupants often prefer light levels that are less than 100%.

Evaluation of Lighting for DR

The great potential for widespread use of lighting for DR will only be realized if more lighting systems are installed or upgraded to have the following features:

1. Centralized controls.
2. Zoning that allows light levels to be reduced with some degree of resolution that is minimally disruptive to building occupants.
3. Flexibility for various end-use scenarios.

SUMMARY AND FUTURE DIRECTIONS

This paper has presented a review of DR control strategies in commercial buildings based on a combination of results from field studies in 30 buildings over a three year period. The field studies have shown that there is a significant opportunity to enable DR capabilities in many existing buildings using existing EMCS and lighting controls. Further research is needed to understand the prevalence of controls in existing buildings to support a broad based deployment of these strategies. Newer, more advanced controls provide greater capability than older systems. Future work in this project will explore the applicability of these strategies to various building types, sizes, and climates.

ACKNOWLEDGMENTS

The authors are grateful for the extensive support from numerous individuals who assisted in this project. Many thanks to the engineers and staff at each building site. Special thanks to Ron Hofmann for his conceptualization of this project and ongoing technical support. Thanks also to Laurie ten Hope, Mark Rawson, and Dave Michel at the California Energy Commission. Thanks also to the Pacific Gas and Electric Company who funded the Automated CPP research. This work described in this report was coordinated by the Demand Response Research Center and funded by the California Energy Commission, Public Interest Energy Research Program, under work for others contract No. 150-99-003, Am #1 and by the U.S. Department of Energy under contract No. DE-AC03-76SF00098.

REFERENCES

1. Akashi, Y.; Neches, J. Detectability and acceptability of illuminance reduction for load shedding, IESNA Annual Conference Proceedings, August 3–6, 2003.
2. Goldman, C.; Hopper, N.; Sezgen, O. et al.; *Does Real-Time Pricing Deliver Demand Response? A Case Study of Niagara Mohawk's Large Customer RTP Tariff LBNL-54974*, 2004.
3. Kiliccote, S.; Piette, M.A. Control technologies and strategies characterizing demand response and energy efficiency. In *Proceedings of the Fifth Annual International Conference on Enhanced Building Operations*, Pittsburgh, PA, October 2005; LBNL #58179.
4. Piette, M.A.; Sezgen, O.; Watson, D.; Motegi, N.; Shockman, C. *Development and evaluation of fully automated demand response in large facilities*, CEC-500-2005-013. LBNL-55085. January 2005a. Available at http://drrc.lbl.gov/drrc-pubs1.html
5. Piette, M.A.; Watson, D.; Motegi, N.; Bourassa, N.; Shockman, C. Findings from the 2004 Fully Automated Demand Response Tests in Large Facilities, CEC-500-03-026. LBNL-58178, September 2005b. Available at http://drrc.lbl.gov/drrc-pubs1.html
6. Xu, P.; Haves, P. Case study of demand shifting with thermal mass in two large commercial buildings. ASHRAE Trans. **2005**, *LBNL-58649*.

BIBLIOGRAPHY

1. Kiliccote, S.; Piette, M.A.; Hansen, D. Advanced control and communication technologies for energy efficiency and demand response. In *Proceedings of Second Carnegie Mellon Conference in Electric Power Systems: Monitoring, Sensing, Software and Its Valuation for the Changing Electric Power Industry*, Pittsburgh, PA, January 2006; LBNL #59337.
2. Piette, M.A.; Watson, D.; Motegi, N.; Kiliccote, S.; Xu, P. *Automated critical peak pricing field tests: program description and results*, Report LBNL-59351, January 2006.
3. Xu, P.; Haves, P.; Piette, M.A.; Braun, J. Peak demand reduction from pre-cooling with zone temperature reset of HVAC in an office. In *Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove, CA, 2004; LBNL-55800.

Demand Response: Load Response Resources and Programs

Larry B. Barrett
Barrett Consulting Associates, Colorado Springs, Colorado, U.S.A.

Abstract
"Demand response" is a relatively new term to the electric utility industry for an old concept called peak load management. Demand response has gained currency, as the historically cumbersome peak load management programs have been transformed by real-time monitoring, digital controls, and robust communications. The costs of managing demand response resources have become more competitive relative to the costs of old load management techniques such as central station power plants and electric transmission system upgrades. Accordingly, demand response has joined the energy management lexicon as a more refined and flexible alternative to the old term of peak load management.

DEFINING DEMAND RESPONSE

Demand response in electricity markets is defined as "…load response called for by others and price response managed by end-use customers."[1] The definition of demand response conveniently divides activities into two categories: load response and price response, also called economic demand response.

Load response occurs when end users react to requests for reducing electric demand. Examples of load response programs are interruptible programs, curtailable programs, and cycling programs.

Price response occurs when end users react to price signals. Examples of these economic demand response programs include time-of-use rates, real-time pricing, and critical peak pricing.

The scope of this entry is focused on load response programs within the broader category of demand response. Material is presented elsewhere in this publication on the economic demand response programs such as real-time pricing.

The essential difference between load response and price response is who initiates and who follows in the short-term. In load response, an energy supplier initiates the call for load management and the customer acts under the terms of some agreement. In price response, the energy supplier sets the rates and the customer is responsible for initiating usage limitation actions, if any.

Keywords: Demand response; Peak load management; Load response; Curtailable load programs; Building automation; Air conditioner cycling; Ancillary services.

BENEFITS OF DEMAND RESPONSE

The potential for demand response is significant. For example, the largest grid operator in the United States, known as the PJM, counted demand response resources as equivalent to 5% of its peak demand in its Mid-Atlantic and Great Lakes regions.[2]

The benefits of demand response are varied and numerous, particularly when considering the many parties such as power grid operators, local electric distribution companies, facility managers, and all the customers who would have to pay higher rates otherwise. The benefits include the following:

- *Power system reliability*—Local and regional electric grids achieve greater reliability while avoiding blackouts and voltage reductions in emergency situations when customers are able to reduce loads on the grid. It has been estimated that "Power interruptions and inadequate power quality already cause economic losses to the nation conservatively estimated at more than \$100 billion a year."[3]
- *Market efficiency*—Costs of power production and distribution can vary dramatically during the course of a day, a week, and a year as some plants operate only for the hours of peak demand. Yet standard rates present constant cost signals to consumers rather than the actual fluctuations in costs. Demand response programs provide incentives for customers that are able to reduce and shift loads, which not only benefit the program participants but all customers. The result is a more efficient use of the power generation, transmission, and distribution systems.
- *Bill savings*—Energy savings achieved during demand response periods typically translate to bill savings and are reinforced when load reductions are further rewarded with incentive payments. Electric bill savings

could reach about $7.5 billion per year in the United States according to a study for the Federal Energy Regulatory Commission (FERC).[4]

- *Cost reduction*—While similar to bill savings, cost reduction is more of a long-term benefit. Demand response reduces the long run costs of new power plants, defers new and expanded high voltage transmission systems, and mitigates the overloading of low voltage distribution systems.
- *Environmental quality*—To the extent that peaking power plants are old and inefficient, their reduced use can improve the environment. Even where peaking power plants are relatively clean in terms of emissions, demand response that results in a net reduction in energy use can improve environmental quality.
- *Customer service*—Demand response provides something that customers want—a choice. With demand response programs, customers that are able and willing to adjust their electricity usage for a few hours have another way to manage energy besides simply using less.
- *Risk management*—Providers of retail electricity must cover the risks of price volatility in wholesale markets if retail prices are less volatile. Demand response programs help cover those risks through greater energy availability, reliability, modularity, and dispensability.
- *Market power mitigation*—Market power refers to concentrating the central generation capacity into a few organizations. Demand response programs with hundreds, and indeed thousands, of owners of distributed assets can be called upon to help mitigate market power.
- *Complements to energy efficiency*—Demand response resources are discontinuous or occasional since they are only called upon for a few hours at a time. Energy efficiency actions are usually more permanent or continuous. Even where energy efficiency investments are made, there is opportunity for demand response participation.

LOAD RESPONSE PROGRAMS

Load response programs include those where the customer is responding to requests for short-term peak load reduction. They are sometimes referred to as "reliability-driven" programs in contrast to "market-based pricing" programs.[4]

Load response programs may be divided into two classes. The first is where virtually the entire facility or operation is interrupted. The second is where part of the facility reduces demand on the power grid so that the load is curtailed.

Interruptible Load Response Programs

Interruptible load response programs operate, as the name suggests, where the customer's entire facility or operations must be interrupted or shut down. A few circuits may be exempted to support lighting, HVAC, communications, and computer services in the administrative portions of the facilities. However, the major proportion of the electrical load is made up of much larger uses of power (e.g., production lines) which may be reduced through power interruptions.

A feature of interruptible programs is the ability of the utility to control the power flow. Thus, a feeder may be opened to prevent power to flow to the participating facility.

Another feature is that participation is voluntary and mandatory at the same time. Customers are free to join interruptible programs; however, once in the program, they must agree to power interruptions.

The incentive has historically been quite attractive to certain types of customers. Typically customers are rewarded with a lower rate that can prove to be a significant discount over the course of a year. In exchange, the customer agrees to interruptions. In most jurisdictions, the customer rarely has power interrupted. Thus, over decades of operation, these customers took advantage of the rate discount without any particular cost or inconvenience. And utilities offered the rate more as an inducement for economic development purposes to lure new facilities to their service territory.

Failure to perform or comply can be expensive. Penalties may be applied to facilities found not to be interrupting. These penalties can be quite substantial and may in fact negate the savings from the year-round rate reductions. Among other features, warning times of no more than 30 min may be imposed before interruption is required.

Participants in interruptible load response programs are typically industrial customers. One reason is that the utilities prefer larger operations with significant load reduction potential. A second is that industrial facilities are more likely to be subject to competing economic development offers, compared to commercial facilities such as office buildings or hospitals.

Load Curtailment Programs

These programs allow facility owners to curtail parts of their operations rather than entire facilities. Another feature of load curtailment programs is that smaller facilities may be eligible since lower thresholds of load reduction potential may still qualify. For example, many program participants must be able to provide only 100 kW of load reduction to qualify.

Another attraction of load curtailment programs is the voluntary nature of participation. Not only is selection into the program voluntary, each curtailment event may be voluntary. That is, when the utility calls for a load reduction, the customer may choose to curtail and get paid for the reduction or to continue operations and pay the price per the agreement. The customer may not be mandated to reduce their load, but failure to meet target load reductions can result in penalties. However, the penalties are not usually as severe as with interruptible load response programs.

Another feature is advance warning. Some programs may provide a 24-h notice before the curtailment event. Others offer a 2-h notice and others still a 30 min notice. Payments to customers for load curtailment may increase as the length of advance notice decreases.

Total power reduction may take place for certain facilities, such as those with standby generators sufficient to carry an entire facility or operation. While load curtailment implies a partial as opposed to total load reduction in most programs, customers may elect to provide the maximum load reduction possible by disconnecting from the grid either figuratively or in fact. In this case, the customer's load is generally met by standby generators designed to carry the entire facility.

Demand Buyback Programs—Pay for Performance

One variation on the operation of load curtailment programs is the demand buyback program. Also known as pay for performance, customers curtail loads in a two-step process.

First, the customer is notified by the energy supplier that a curtailment event is likely and bids will be accepted for peak load reduction. The customer decides whether to exercise the option of offering a certain amount of load reduction. The customer may also be required to suggest the price it wants to be paid in order to participate.

Then, the energy supplier has the option to accept the offer from the customer. Once accepted, the customer is typically obliged to meet the load reduction target in exchange for the customer's requested incentive.

The notice that a customer may receive before the possible curtailment event may vary from an hour to a couple of days. Once the customer bids and the energy supplier accepts the bid, the transaction should go through. If the customer does not shed load, a penalty may be imposed by the utility. The size of the penalty may be determined by various factors: the penalty may be based on just the load reduction that failed to materialize, or it could be based on the total load reduction promised under the buyback arrangement.

The amount of the buyback incentive may vary from event to event. Similarly the penalty may vary by event. The incentive may be based on kilowatt-hour reductions, kilowatt reductions or some other measure of performance. The incentive amount plus the performance achieved should determine the payment for customer participation.

COMMERCIAL BUILDINGS AND DEMAND RESPONSE SOLUTIONS

There are many ways to operate facilities for peak demand management.[5] Many of these solutions not only reduce peak demand but also save energy. Of course, any combination of solutions will vary by geography as well as from facility to facility and industry to industry.

Before participating in a demand response program, it is helpful for a facility to conduct an audit of the potential for peak load reduction. The utility that sponsors the demand response program may even provide an audit at no charge to inventory the facilities and equipment for their potential load reduction. An added benefit may be that in addition to obtaining a report on its peak load management opportunities, the audit may also suggest how the customer could increase general savings from energy efficiency upgrades.

The asset management options or resource solutions for demand response may be divided into practices requiring little investment and measures defined by significant investments. The asset options may also be divided according to energy end-uses as presented below.

Lighting represents a significant opportunity for load reduction. During peak periods, turn off unnecessary lights, including storerooms, mechanical rooms, wall washers, and spotlights. For retail facilities, this includes display lights on low-value or on-sale merchandise. In office settings, turn off lobby lights and a portion of hallway lights. In all cases, its advised to turn off exterior lights that happen to be on during the day. If the facility is wired for demand reduction potential, turn off selected lights on circuits with separate controls. For example, some fixtures allow users to turn off half the lights in a luminaire or turn off some fixtures while others continue to operate. For example, some circuits allow one row of fixtures to be turned off, while the adjacent row operates. Other lighting reduction options for peak periods include turning off lighting next to windows. Alternatively, dim lights where turning off lights is not feasible.

Air conditioning represents another large opportunity for reducing electricity usage. Options to reduce peak demand include increasing temperature set points on air conditioning thermostats and relying on outside air for cooling using economizers when weather conditions permit. With advance notice of a few hours or a day before an event, there is time to pre-cool space below normal temperatures prior to peak load conditions. Also, rotate the operation of chillers and packaged rooftop units and turn off condensing units while maintaining fan operations to continue air movement. If there are two-speed compressors on rooftop units, use the lower speed. Institute soft start procedures for multiple air conditioning units and let them coast during peak periods. Thermal energy storage is an air conditioning option that may be worth the investment, particularly in new facilities. This allows air conditioning to be supplied from chilled water or ice stored in tanks during peak hours. All of the air conditioning load can be moved off-peak with large thermal energy storage systems. Even partial systems allow substantial loads to be moved off peak.

Heating systems offer some potential as well for alleviating peaking electric systems in the winter. Options include reducing temperature set points on heating systems and pre-heating space above normal temperatures prior to peak load conditions. Electric thermal energy storage systems allow for off-peak charging at night with heat provided during peak daytime hours.

Ventilation options include installing carbon dioxide sensors that allow air intake to be reduced during peak hours, when levels of indoor air quality are acceptable. Reduced ventilation with outdoor air means less air conditioning load and reduced use of space conditioning systems. Separate carbon monoxide sensors for garages may prevent the operation of supply and exhaust ventilation systems that operate continuously, even though traffic patterns may only warrant operation for a few hours each day. Where permitted by codes, the carbon monoxide sensors can be tied to energy management systems and significantly reduce fan operation during peak load hours with little traffic.

Finally, building automation systems allow peak load strategies to be introduced reliably and consistently for the aforementioned air conditioning, heating, ventilation, and lighting systems. Automated controls also allow for the consolidation of multiple sites on a real-time basis to enable a single point of operation.

Options for commercial refrigeration including turning off some units for a few hours, postponing defrost cycles, staging operations to gain load diversity, and turning off electric strip heaters designed to remove moisture from glass covers. Water heaters are excellent candidates for peak load management, where storage capacity allows a facility to cycle units off for several hours.

Any facility with standby generators for emergency operation may be a candidate for operation during economic peak load management events. If cost-effective for such economic load management programs, facility owners may want to make upgrades to operate fully powered by the generators while in parallel to the electric grid. Another generic strategy for peak load reduction is to take some of the elevators and down escalators out of service for a few hours each day.

INDUSTRIAL DEMAND RESPONSE SOLUTIONS

Industrial facilities also present numerous opportunities to reduce and shift loads for several hours at a time. Some facilities can act within seconds of notification and most within the half hour. However, many facilities prefer advance notice of 2 h or a full day. The greater the advance notice the more consumers can do to reduce electric loads.

Prime candidates for demand response are facilities with storage capabilities in terms of their production materials and product shipping inventories. Air separation plants that produce oxygen, nitrogen, and other gases are examples where products can be stored and inventories managed. These plants operate automatically, and often remotely, allowing few complications with labor and other management considerations to accommodate requests for peak load reduction.

Water storage operations provide other options to reduce peak load. Manufacturing facilities with significant demands for water can coast through load curtailments by relying on stored water. Or, storage tanks may be filled at a slower rate during curtailment events, allowing some net drawdown of water supplies, and then the tanks can be replenished to capacity during off-peak conditions.

Public water delivery systems present a complex mix of pumping, gravity, and storage. With proper configuration, the pumping systems can be shut down or slowed to reduce peak loads. In wastewater treatment plants with multiple aerators, several options are available. One is to turn off the aerators for a few hours. Another is to cycle them so not all units are running at the same time. Still another option is to slow the aerators operation with variable frequency drives on the motors.

Refrigeration systems can assist in reducing peak loads with sufficient notice. A day's advance notice allows operators to pre-cool refrigeration cases to a lower than normal level and then coast through the curtailment period. Or, refrigeration temperatures may be permitted to migrate upward by a few degrees during curtailments and be restored afterwards.

Batch processing operations also lend themselves to curtailment. Once a batch is completed, the process may be halted and the load reduced until the curtailment event is over. In some processes, batches may be interrupted and then restarted without complications.

Continuous processing operations are also candidates for peak load management if they can be slowed down. Plants with variable frequency drives on pumps and motors can slow their operations for a few hours and still maintain some production.

Compressors are a large consumer of energy and present many options to save energy and peak demand. Plants with sufficient compressed air storage may be able to turn off the compressors for short periods of time. Plants with multiple compressors may be able to rotate their operation and reduce peak demand from simultaneous operation.[6]

Industrial processes with standby generators may not need to change their production operations at all. Generators may pick up significant portions of the internal plant load and free up capacity on the power grid for other purposes.

RESIDENTIAL DEMAND RESPONSE SOLUTIONS

Residential programs for demand response are in place with tens of thousands, and in some utilities, hundreds of thousands of homes participating. Three resources may be targeted for peak load reduction: central air conditioners, electric water heaters, and swimming pool pumps.

The large majority of demand response reduction comes from central air conditioners. While an average load reduction of 1 kW per residential air conditioner during peak summer periods may seem small, when aggregated over numerous households the available load response resource is significant.

Improved technologies are driving the growth of demand response with residential appliances such as air conditioners. In the 1980s, control switches were typically installed on outdoor condensing units of air conditioning systems. The switches were conspicuous and subject to damage. Now, special controls can be installed internally and linked with smart thermostats to cycle climate control systems. Also, new systems allow customers to override the controls on an exception basis and therefore increase load reduction participation rates.

Another improvement in load cycling has been with communications systems. Legacy products received signals to interrupt operations without an ability to acknowledge the request. Now, two-way communications are possible, allowing for more sophisticated control strategies and incentive plans to encourage wider participation.

Another improvement is in the measurement and verification of electrical usage. With better communications, it is possible to determine if systems are being cycled and for how long. Also, improved database systems and management software foster more robust and dependable measurement and verification protocols.

New technologies also allow alternative programs including different cycling strategies that can be varied according to a participant's energy settings and energy use. When controls are sensitive to natural duty cycles there is greater reliability in achieving optimum reductions. Also, there are more alternatives for different types of incentive plans such as those based on the number of cycling events, override frequencies, temperature adjustments, and the temporary shutdown of the compressor system for several hours under 100% cycling.

Load management strategies range from 25% cycling, where units are cycled off only seven and a half minutes per 30-min period, to 100% cycling where the compressor may be turned off for the duration of the event. Cycling can also be exercised from less than 2 h to over 6 h per day. A risk of cycling too short in hours and too low in percentage of time off is that the natural duty cycle is being replicated with little appreciable savings in peak demand.

An important advantage of load cycling is the ability to target certain neighborhoods and areas within a utility's service territory. This can be particularly beneficial where specific parts of the electrical system are in danger of being overloaded from rapid growth or aging infrastructure.

METERING, COMMUNICATIONS, AND CONTROL

Metering

Extra metering is typically required for commercial and industrial facilities. The most common application is with interval meters. These meters record usage at least hourly and for many applications, on 15 min intervals. As demand response expands into more time-sensitive operations, 5-min intervals may become more common. Interval meters are also called smart, automated, or advanced meters.

Interval meters record electricity consumption in kilowatt-hours, while demand is recorded in kilowatts and time of use. Some meters calculate the maximum or peak demand over a specified time period such as a day, week, or month. When connected to a personal computer with a modem, interval meters offer the utility of a rich database to analyze energy use levels and patterns.

Interval meters are essential to estimating the load reductions achieved by demand response programs. Load profiles may be developed for each time interval on the designated curtailment days. The load profiles can then be compared with normal days to calculate load reductions on which to base payments.

Communications

Communicating and advance notice of curtailment periods are becoming more sophisticated. In the early days of load response, communications were made manually through a telephone call or by an automated notification arrangement over dedicated telephone lines. Today, there is a trend toward wireless communications where signals are quite economical in short bursts of airtime.

Often multiple forms of communication are employed to insure that facility managers and operators receive notice of planned curtailments. Thus, a request for curtailment may be communicated by some combination of FAX, telephone, email, and pager.

Two-way communications are another feature of many demand response programs. The end-use facility is configured to receive a signal and request for curtailment. In addition, communications are sent back acknowledging the receipt of the request and, just as important, the real-time load levels. Such communications based on the interval meter recordings and stored on a web-enabled personal computer may be sent to the utility, the grid operation, or an additional third party that monitors performance.

The availability of load performance is an attractive feature of demand response programs. This performance information may be of value to the corporate energy engineer, the store manager, the plant superintendent, and others in the customer organization. A facility's finance and accounting departments may find the information of value, and, if participation may affect shipping schedules and sales, even those in marketing and sales.

Control

With data comes information, and with information comes control. Data from interval meters may be integrated with building management systems and energy management systems. These systems can control lighting, thermostat settings, equipment cycling, and other operations. Certain systems may be programmed to recognize a curtailment request and automatically shift into a different operating protocol to accommodate the event.

Residential applications are amenable to more sophisticated controls for load response. Programmable thermostats can be enhanced to accept signals for curtailment or cycling. Some thermostat models may also be configured to adjust the set-point by some specified amount, such as four degrees, with the effect of reducing air conditioning usage during the curtailment event.

Metering and communications are important in another way, namely by applying credit for a customer's load response performance. Credit may be issued in the form of a check or electric bill reduction. Done manually, settlement or payment for load response can take months where multiple parties are involved in requesting and managing the load reductions. Advances in metering and communications help accelerate the settlement process.

ANCILLARY SERVICES

Ancillary services refer to such functions as regulation, spinning reserve, supplemental reserve, and replacement reserve on the bulk power grid.[7] Regulation services operate in fractions of a second to maintain the balance between power generation and customer loads while still maintaining voltages within required ranges.

Spinning reserve services are called upon in the event of a generator outage or transmission interruption. Spinning reserves need to be synchronized to the grid and meet capacity within 15 min. Supplemental reserves are similar to spinning reserves but do not need to be synchronized immediately, as long as they can reach capacity within 15 min. Replacement reserves are similar to supplemental reserves but have a 30–60 min window to reach capacity. The supplemental and replacement reserves may be called upon to replace spinning reserves, allowing the spinning reserves to stand down and be ready for another contingency.

Traditionally, ancillary services have been provided by central generation plants. Plants are kept in spinning reserve, but not under load, in the event that there is a system failure on the grid and additional load is needed suddenly. The reserves must be able to supply capacity and regulate power within minutes.

However, many demand response resources perform within the short time deadlines needed to supply spinning reserves. Some demand response resources can respond within seconds. For example, standby generators can respond within 10–15 s.

As another example, air conditioner cycling programs for residential and small commercial buildings can be signaled within a minute. If called upon for spinning reserves, cycling programs are even more advantageous, since they can easily operate for the minimum of 30 min required by ancillary services.

When air conditioner loads are interrupted to meet needs for spinning reserves, the available capacity can be triple the amount of capacity that is available compared to cycling.[8] The reason is that cycling disrupts, but does not necessarily eliminate, the natural diversity of air conditioner operations over each 30 min cycle. Interrupting loads for a spinning reserve can prevent even this part-time operation over 30 min.

The 30 min time period allows central station generation resources to come on line and make up for the capacity lost in spinning reserves. Then, the cycling schedule can be discontinued and the resources once again are available in standby to provide spinning reserves. Multiple parties are satisfied since power plants can operate for hours to meet system needs, while if cycling programs operate for too many hours, the customers may start to experience more discomfort and inconvenience than they are not likely to notice in a 30 min event.

Another benefit is that the aggregation of many demand response resources, while small individually, makes it highly probable that the assigned ancillary services will be achieved in total. This means that demand response resources are more reliable, when compared to a central generation resource, where the failure of one turbine could cause large losses of spinning reserves.

Compared to central station power plants, demand response resources can help "level the playing field" by providing an equivalent capacity, with high reliability, in an economic manner.

ECONOMICS OF DEMAND RESPONSE

By balancing financial benefits with costs, the primary motivation for end users to manage peak loads is economic. Whether the customer is residential, commercial, or industrial, incentives must be sufficient to cover whatever costs, inconvenience, and perhaps discomfort that may arise with load reductions. Furthermore, electric bill savings may be realized through reduced usage.

Another motivation to participate in demand response programs is for community reasons such as helping to improve the environment. A third motivation is to gain information about energy use and operating conditions associated with the more detailed monitoring protocols attendant to demand response programs.

Incentives may be paid from multiple parties. Grid operations are willing to pay for demand response resources at many times the normal electric rates. For example, in 2004, the New York Independent System Operator and the Independent System Operator of New England offered $500 per megawatt-hour or 50 cents per kilowatt-hour for load response resources during peak hours on the grid.

Other parties that may pay for demand response resources are traditional utilities with operations that are vertically integrated from the power plant to the meter, or utilities with only distribution businesses attempting to reduce high demand charges. There can be third parties such as curtailment service providers that make a business of aggregating customer loads to bid into demand response programs.

A key consideration in the economics of demand response is the cost of metering, communication, and control. In some cases, the cost may be underwritten by utilities, grid operators, and even government agencies with energy responsibilities.

There may be other costs to consider, such as modifications to production equipment to increase capacity or improve controls. Of course, labor costs associated with interrupting operations are a factor, particularly if overtime wages must be paid to accommodate the shifts in production schedules.

In general, it is advantageous to anticipate demand response opportunities when building new facilities or upgrading load capacity and operations. As demand response programs continue to expand in size and scope, the relationship of benefits and costs for participation should continue to improve.

CONCLUSION

Load response programs, as one of the principal forms of demand response, are an important way to achieve peak load management in electricity markets. There are many benefits including increased reliability of the power grid, higher efficiency of energy markets, and increased savings in consumer energy bills. Load response can be achieved in commercial, industrial, and residential buildings from numerous assets including air conditioning, water heating, and refrigeration. Standby generation equipment is a common asset deployed for load response in commercial and industrial facilities.

Load response programs are enabled by improved technologies in metering, communications, and control. The economics of load response favor utility systems with high costs, such as those associated with capacity shortages, operating inefficient units at peak loads, and transmission bottlenecks. Financial rewards from load response are expanding beyond the traditional markets for generation and transmission capacity shortages that may be anticipated hours or days in advance. Load response is gaining acceptance in ancillary services markets to provide stability on the power grid with only a few seconds or minutes notice.

REFERENCES

1. Peak Load Management Alliance. Demand Response: Principles of Regulatory Guidance. February 2002, 1.
2. Barrett, L.B. Demand Response Resources Ready and Able, *Barrett Consulting Associates Report*, January, 2005.
3. EPRI Technology action plan addresses western power crisis. EPRI J. **2001**, 5.
4. U.S. Government Accounting Office. Electricity Markets: Consumers Could Benefit from Demand Programs, but Challenges Remain. August 2004, 4.
5. California Energy Commission. Enhanced Automation: Technical Options Guidebook. 2002.
6. U.S. Department of Energy, Office of Industrial Technologies. Best Practices: Compressed Air System Renovation Project Improves Product at a Food Processing Facility. June 2001.
7. Hirst, Eric. Price-Response Demand as Reliability Resources. Oak Ridge, Tennessee. April 2002, 7.
8. Kirby, B.J. Spinning Reserve from Response Loads. Oak Ridge National Laboratory, ORNL/TM 0-19, March 2003, 35.

Demand-Side Management Programs

Clark W. Gellings
Electric Power Research Institute (EPRI), Palo Alto, California, U.S.A.

Abstract
Demand-side management is the active planning and implementation of programs that will change consumers' use of electricity. These programs may encourage the adoption of more efficient appliances, the use of new technologies, and the time these and other devices are used.

INTRODUCTION

Demand-side management programs result from the planning and implementation of those activities designed to influence consumer use of energy in ways that will produce desired changes in the time pattern and magnitude of energy demand. Programs and initiatives falling under the umbrella of demand-side management include load management, new uses, strategic conservation, electrification, and adjustments in the market share of energy-consuming devices and appliances.

Demand-side management includes only those activities that involve a deliberate intervention in the marketplace. This intervention has often been affected by electric or natural gas utilities. Examples of demand-side management programs include those that encourage consumers to install energy-efficient refrigerators, through either incentives or advertising.

Demand-side management extends beyond conservation and load management to include programs designed specifically to modify energy use in all periods. Thus, demand-side management alternatives warrant consideration by entities such as energy and energy service suppliers with ambitious construction programs, those with high reserve margins, and those facing high marginal costs.

Key definitions for demand-side management are given below.

WHY CONSIDER DEMAND-SIDE MANAGEMENT

Since the early 1970s, economic, political, social, technological, and resource supply factors have combined to change the electricity sector's outlook for the future. Many utilities face significant fluctuations in demand and energy growth rates, declining financial performance, and political or regulatory and consumer concern about rising prices. Although demand-side management is not a cure-all for these difficulties, it does provide additional alternatives.

For utilities facing strong load growth, load management and strategic conservation can provide an effective means to reduce or postpone acquisition of power contracts or construction of new generating facilities; for others, load growth can improve the utility load characteristics and optimize asset utilization. Changing the purchase pattern or the amount of consumer demand on an energy system can reduce operating costs.

Implementing demand-side management can lead to greater flexibility in facing rapid change in today's business environment. Although demand-side management will not solve all the problems facing the energy industry and its stakeholders, it does provide additional alternatives for meeting the challenges of the future (Fig. 1).

SELECTING THE RIGHT ALTERNATIVE

Demand-side management encompasses planning, evaluation, implementation, and monitoring of activities selected from among a wide variety of programmatic and technical alternatives. Due to the large number of alternatives, assessing which alternative is best suited is not a trivial task. The choice is complicated by the fact that the attractiveness of alternatives is influenced strongly by specific local and regional factors, such as industry structure, generating mix, expected load growth, capacity expansion plans, load factor, load shapes for average and extreme days, regulatory climate, and reserve margins. Therefore, it is often inappropriate to transfer these varying specific factors from one area to another without appropriate adjustments. In addition, the success of any alternative depends on the specific combination of promotional activities selected to aid in promoting customer acceptance.

Keywords: Demand-side management; Energy efficiency; Conservation; Load management; Consumer; Load shape; Load control; Marketing; End use.

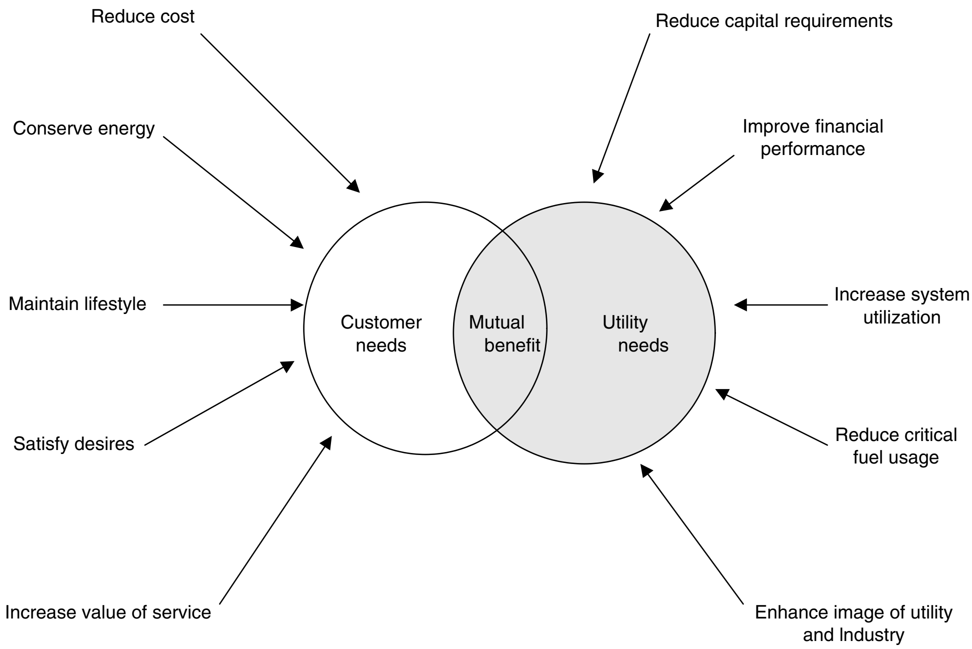


Fig. 1 Goals of demand-side management programs.

LOAD-SHAPE CHANGES

Although there is an infinite combination of load-shape changing possibilities, six can illustrate the range of possibilities: peak clipping, valley filling, load shifting, strategic conservation, strategic load growth, and flexible load shape. These six are not mutually exclusive and may frequently be employed in combinations.

The demand-side management planning approach does provide policymakers with a whole new set of alternatives with which to meet energy needs. The concept that the load shape is not fixed but can be altered deliberately opens a new dimension in planning and operation.

DEMAND-SIDE MANAGEMENT PRACTICE

Demand-side management can help achieve a broad range of operational objectives merely by changing the system's load shape. Numerous industries have found that changing the pattern of the demand for their product can be profitable. Telephone utilities, for example, have long offered reduced evening rates to shift demand and to encourage off-peak use. Airlines offer night coach fares, and movie theaters offer reduced matinee prices—examples of deliberate attempts to change the demand pattern for a product or service.

In an electric utility, the physical plant is designed to serve the projected demand for electricity in the least-cost manner, given a specified level of desired quality and reliability. If the load shape is not fixed but may be altered, the cost of serving the load can be reduced still further. Cost reductions due to changes in the load shape arise primarily from three attributes:

- Reduction in the requirements for new assets or energy
- Higher utilization of facilities
- More efficient operation of facilities

System expansion can be delayed or eliminated and the use of critical resources reduced by significantly reducing energy and peak consumption through the use of demand-side management.

Higher utilization of existing and planned facilities can also be achieved through a program of load growth. Although such programs increase total costs due to higher fuel and other operating expenses, they reduce unit costs by spreading the fixed costs (debt service and dividends) over more units of energy sold.

Six generic load-shape objectives can be considered to be part of demand-side management (Fig. 2).

SELECTING DEMAND-SIDE MANAGEMENT ACTIVITIES

Although customers and suppliers act independently, resulting in a pattern of demand, the concept of demand-side management implies a supplier/customer relationship that produces mutually beneficial results. To achieve these mutual benefits, suppliers must carefully consider the manner in which the activity will affect the patterns and amount of demand (load shape). In addition, suppliers must assess the methods available for obtaining customer

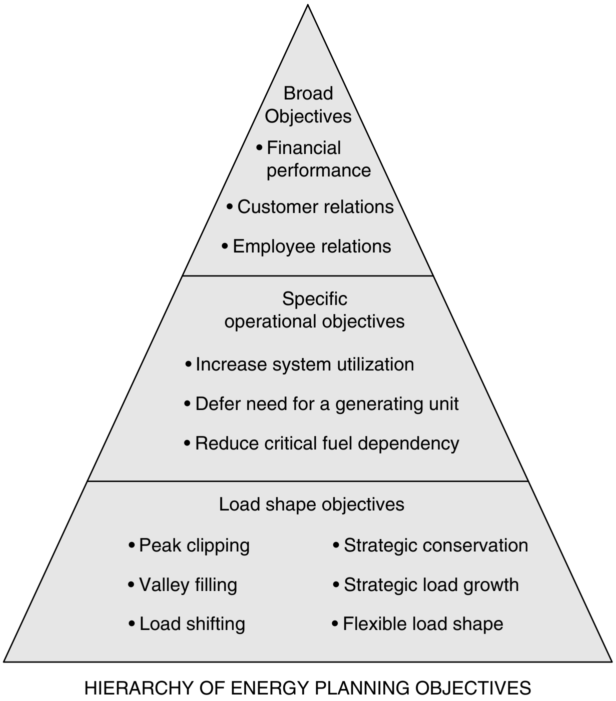


Fig. 2 Hierarchy of energy planning objectives.

participation, and the likely magnitudes of costs and benefits to both supplier and customer prior to attempting implementation.

Because there are so many demand-side management alternatives, the process of identifying potential candidates can best be carried out by considering key aspects of the alternatives in an orderly fashion. Demand-side management activities can be categorized in a two-step process:

- Step 1: load-shape objectives
- Step 2: end-use technology activities

LOAD-SHAPE OBJECTIVES

The first step in identifying demand-side alternatives is the selection of a load-shape objective to ensure that the desired result is consistent with goals and constraints.

END USE

When the load-shape objective has been established, it is necessary to find end uses that can achieve it. This identification process involves three dimensions. The first dimension involves identifying the appropriate end uses whose consumption characteristics generally match the requirements of the load-shape objectives. In general, each end use exhibits predictable demand or load patterns.

Nine major residential end-uses of electricity are examples of having the most potential for electric demand-side management. They are space heating, space cooling, water heating, lighting, refrigeration, cooking, laundry, swimming pools, and miscellaneous other uses. Each of these end uses provides a different set of opportunities to meet electric load-shape modification objectives. Some of the end uses can serve successfully as the focus of programs to meet any of the load-shape objectives; others realistically can be useful for meeting only one or two objectives. In general, space heating, space cooling, and water heating are the residential end uses with the greatest potential applicability for achieving objectives. These end uses tend to be among the most energy intensive and among the most adaptable.

TECHNOLOGY ALTERNATIVES

The second dimension of demand-side management alternatives involves choosing appropriate technology alternatives for each end use. This process should consider the suitability of the technology.

Residential demand-side management technologies can be grouped into four general categories:

- Building-envelope technologies
- Efficient equipment and appliances
- Thermal storage technologies
- Energy and demand control technologies.

These four main categories cover most of the available residential options. Many of the individual options can be considered to be components of an overall program.

HOW TO SELECT ALTERNATIVES

Selection of the most appropriate demand-side management alternatives is the most crucial question. The relative attractiveness of alternatives depends on specific characteristics, such as load shape, summer and winter peaks, generation or product system mix, customer mix, and the projected rate of load growth.

MARKET IMPLEMENTATION METHODS

Among the most important dimensions in the characterization of demand-side management alternatives is the selection of the appropriate market implementation methods (Fig. 3).

Planners and policymakers can choose among a wide range of methods designed to influence customer adoption, which can be broadly classified in six categories:

- Customer education
- Direct customer contact
- Trade-ally cooperation
- Advertising and promotion
- Alternative pricing
- Direct incentives

Energy suppliers, utilities, and government entities have used many of these strategies successfully. Typically, multiple marketing methods are used to promote demand-side management programs. The selection of the individual market implementation method or mix of methods depends on a number of factors, including:

- Experience with similar programs
- Existing market penetration
- The receptivity of policymakers and regulatory authorities
- The estimated program benefits and costs to suppliers and customers
- Stage of buyer readiness
- Barriers to implementation

The objectives of deploying market implementation methods are to influence the marketplace and to change customer behavior. The key question is the selection of the market implementation method(s) to obtain the desired customer acceptance and response. Customer acceptance refers to customer willingness to participate in an implementation program, customer decisions to adopt the desired technology, and behavior change as encouraged by the supplier. Customer response is the actual load-shape change that results from customer action, combined with the characteristics of the devices.

Customer acceptance and responses are influenced by the demographic characteristics of the customer, income, knowledge, and awareness of the technologies and programs available, as well as attitudes and motivations. Customer acceptance and response are also influenced by

Dem–Dry

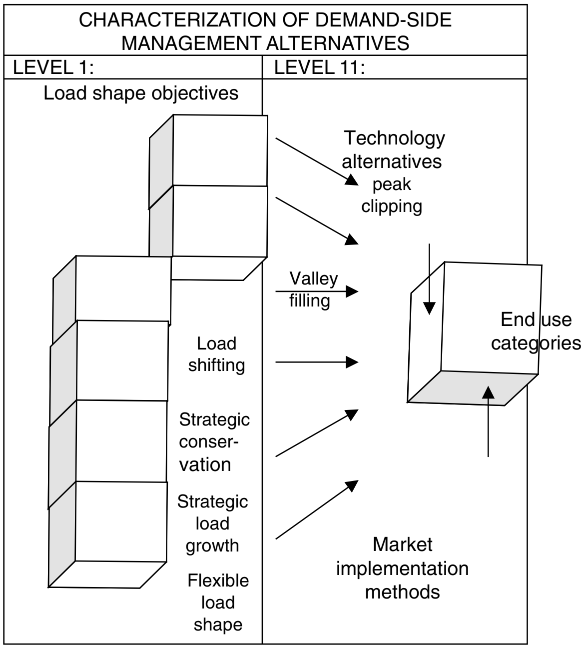


Fig. 3 Characterization of demand-side management alternatives.

other external factors, such as economic conditions, prices, technology, regulation, and tax credits.

CUSTOMER EDUCATION

Many energy suppliers and governments have relied on some form of customer education to promote general customer awareness of demand-side management programs. Brochures, bill inserts, information packets, clearinghouses, educational curricula, and direct mailings are widely used. Customer education is the most basic of the market implementation methods available and can be used to:

- Inform customers about products/services being offered and their benefits, and influence customer decisions to participate in a program
- Increase the perceived value of service to customers
- Inform customers of the eligibility requirements for program participation
- Increase customers' knowledge of factors influencing energy purchase decisions
- Provide customers other information of general interest
- Generally improve customer relations.

DIRECT CUSTOMER CONTACT

Direct customer contact techniques refer to face-to-face communication between the customer and an energy supplier to encourage greater customer acceptance of programs. Energy suppliers have for some time employed marketing and customer service representatives to provide advice on appliance choice and operation, sizing of heating/cooling systems, lighting design, and even home economics.

- Energy audits are particularly useful for identifying heating/air conditioning system improvements, building-envelope improvements, water heating improvements, and the applicability of renewable resource measures.
- Program services involve activities undertaken to support specific demand-side management measures, including heat pumps, weatherization, and renewable energy resources. Examples of such programs include equipment servicing and analyses of customer options.
- Storefronts are business areas where energy information is made available, and appliances and devices are displayed to citizens and consumers.
- Workshops and energy clinics are special sessions that may cover a variety of topics, including home energy conservation, third-party financing, energy-efficient appliances, and other demand-side technologies.
- Exhibits and displays are useful for large public showings, including conferences, fairs, and large showrooms.

TRADE-ALLY COOPERATION

Trade-ally cooperation can contribute significantly to the success of many demand-side management programs. A trade ally is defined as any organization that can influence the transactions between the supplier and its customers. Key trade-ally groups include home builders and contractors, local chapters of professional societies (e.g., the U.S. American Society of Heating, Refrigeration and Air Conditioning Engineers; the Illuminating Engineering Society of North America; and the Institute of Electrical and Electronic Engineers), trade associations (e.g., local plumbing and electrical contractor associations), and associations representing wholesalers and retailers of appliances and energy-consuming devices.

Depending on the type of trade-ally organization, a wide range of services is performed, including:

- Development of standards and procedures
- Technology transfer
- Training
- Certification
- Marketing/sales
- Installation, maintenance, and repair

In performing these diverse services, trade allies may significantly influence the consumer's technology choice. Trade allies can assist substantially in developing and implementing demand-side management programs.

ADVERTISING AND PROMOTION

Energy suppliers have used a variety of advertising and promotional techniques to influence customers. Advertising uses various media to communicate a message to consumers so as to persuade them. Advertising media applicable to demand-side management programs include radio, television, magazines, newspapers, outdoor advertising, and point-of-purchase advertising. Promotion usually includes activities to support advertising, such as press releases, displays, demonstrations, coupons, and contests/awards. Some prefer the use of newspapers, based on consumer research that found this medium to be the major source of customer awareness; others have found radio and television advertising to be more effective.

Advertising and promotion also have widespread applicability. A number of radio and television spots have been developed to promote demand-side management measures. Other promotional techniques used have been awards, energy-efficient logos, and residential home energy rating systems, to name a few.

ALTERNATIVE PRICING

Pricing of energy as a market-influencing factor generally performs three functions:

- Transfers to producers and consumers information regarding the implied value of products and services being provided
- Provides incentives to use the most efficient technologies
- Allocates supply and demand

Alternative pricing through innovative schemes can be an important implementation technique for utilities promoting demand-side management options. Rate incentives for encouraging specific patterns of utilization of electricity can often be combined with other strategies (e.g., direct incentives) to achieve electric utility demand-side management goals.

Various pricing structures are more suited to certain types of demand-side management options. For utilities, time-of-use rates may generally be offered or tied to specific technologies (e.g., storage heating and cooling, or off-peak water heating). They can be useful for thermal storage, energy and demand control, and some efficient equipment options.

DIRECT INCENTIVES

Direct incentives are used to increase market penetration of an option by reducing the net cash outlay required for equipment purchase. Incentives also reduce customer resistance to options without proven performance histories.

The individual categories of direct incentives include

- Cash grants
- Rebates
- Buyback programs
- Billing credits
- Low-interest or no-interest loans

CONCLUSION

Demand-side management offers key advantages to conventional alternatives of supply. Demand-side management is a viable and cost-effective resource that should be part of all electricity supply portfolio planning.

BIBLIOGRAPHY

1. *Customer's Attitudes and Customers' Response to Load Management, Electric Utility Rate Design Study*, Report No. EPRI SIA82-419-6; Published by the Electric Power Research Institute, December 1983.
2. Decision Focus, Inc. *Cost/Benefits Analysis of Demand-Side Planning Alternatives*, EPRI EURDS 94 (RP 1613); Published by Electric Power Research Institute, October 1983.
3. Decision Focus, Inc. *Demand-Side Planning Cost/Benefit Analysis*, Report No. EPRI RDS 94 (RP 1613); Published by the Electric Power Research Institute, November 1983.
4. Decision Focus, Inc. *Integrated Analysis of Load Shapes and Energy Storage*, Report No. EA-970 (RP 1108); Published by the Electric Power Research Institute, March 1979.
5. Demand-Side Management Vol. 1: *Overview of Key Issues*, EPRI EA/EM-3597, Vol. 1 Project 2381-4 Final Report.
6. Demand-Side Management Vol. 2: *Evaluation of Alternatives*, EPRI EA/EM-3597, Vol. 2 Project 2381-4 Final Report.
7. Demand-Side Management Vol. 3: *Technology Alternatives and Market Implementation Methods*, EPRI EA/EM-3597, Vol. 3 Project 2381-4 Final Report.
8. Demand-Side Management Vol. 4: *Commercial Markets and Programs*, EPRI EA/EM-3597, Vol. 4 Project 2381-4 Final Report.
9. EPRI Project, RP 2547, *Consumer Selection of End-Use Devices and Systems.*
10. EPRI Reports prepared by Synergic Resources Corporation. Electric Utility Conservation Programs: *Assessment of Implementation Experience* (RP 2050-11) and 1983 *Survey of Utility End-Use Projects* (EPRI Report No. EM 3529).

Desiccant Dehumidification: Case Study

Michael K. West
Building Systems Scientists, Advantek Consulting, Inc., Melbourne, Florida, U.S.A.

Glenn C. Haynes
RLW Analytics, Middletown, Connecticut, U.S.A.

Abstract

Desiccant dehumidification is primarily a nonresidential end-use technology that can be important to certain commercial businesses such as restaurants, hotels, grocery stores, and hospitals; in public buildings such as courthouses, jails, and auditoriums; and in manufacturing sectors such as pharmaceuticals and microelectronics. This rigorous case study presents results of the field test and performance evaluation of a typical, commercially available, two-wheel gas-fired desiccant air conditioner. Field-measured performance is compared with the manufacturer's specifications, with predictions made using DOE-2 hourly modeling, with all-electric technologies, and with the theoretical limits of the technology. Comparisons were made between the manufacturer's published data, the manufacturer's site test data taken at the time of installation, the collected field data, the computer model, and the theoretical best-case performance. The desiccant unit as installed delivers less cooling and dehumidification capacity than the manufacturer's rating, and much less than it would if the equipment design were optimized and the installation were commissioned. While the measured energy efficiency at peak load conditions is better than the rated efficiency, the data clearly show this rating is not representative of long-term field performance.

BACKGROUND

A field test was initiated to demonstrate and evaluate natural gas desiccant technology in the commercial market segment as a means of controlling weather-sensitive kilowatt electric demand. The serving Florida investor-owned utility, in cooperation with the local gas company, randomly selected and then recruited a commercial customer. The customer installed a new gas desiccant dehumidification system as an alternate technology to the existing electric-DX overcooling and electric reheating system. Advantek Consulting, Inc. was tasked, as an independent third party, with collecting and analyzing field performance data in light of the manufacturer's published data and the results of computer modeling. The customer paid for purchase and installation.

FIELD TESTING

The dehumidification equipment, as well as key components of the building's heating, ventilation, and air-conditioning (HVAC) system, was fitted with a comprehensive instrumentation package to continuously monitor both overall system and sublevel component performance. The field monitoring system collects averaged 1-hour interval data for 35 data points. The most current set of data includes electric kilowatt-hour and natural gas cubic feet (CF) consumption as well as ambient, space, and system temperatures and humidities. The customer integrated the operation of the unit into the existing building management system and is responsible for all maintenance and repairs.

The collected data was screened and used in the calculation of secondary quantities such as the amount of dehumidification capacity delivered, the quantity of moisture removed from the air, and the energy efficiency of the equipment. These quantities were used to assess the performance of the unit as compared with the manufacturer's published performance data. The manufacturer's rated cooling capacity at the peak load condition is 248 MBH (MBH = 1000 Btuh = 0.083 tn); however, the average as-installed capacity was measured to be considerably lower at 155 MBH.

The manufacturer's rated efficiency at the peak load condition (93°F dry bulb, or 78°F wet bulb) is COP 0.73 [COP = (Btuh Capacity)/(Btuh Gas and Electric Input)]; the measured efficiency at this condition was COP 0.83. However, the average as-installed efficiency was measured to be considerably lower than the rated efficiency at COP 0.53. The cooling capacity at peak load was measured to be 19% less than the manufacturer's rating. The heat input at peak load was measured to be 12% less than rated. In comparison, the optimized efficiency of this type of equipment is much higher at COP 1.0–1.2 (Fig. 1).

As designed and installed, the gas dehumidification unit is not optimized nor does it represent the maximum

Keywords: Dehumidification; Humidity; Desiccant; Evaporative; Regeneration; Reactivation; Process; COP (Coefficient of Performance); Preconditioner.

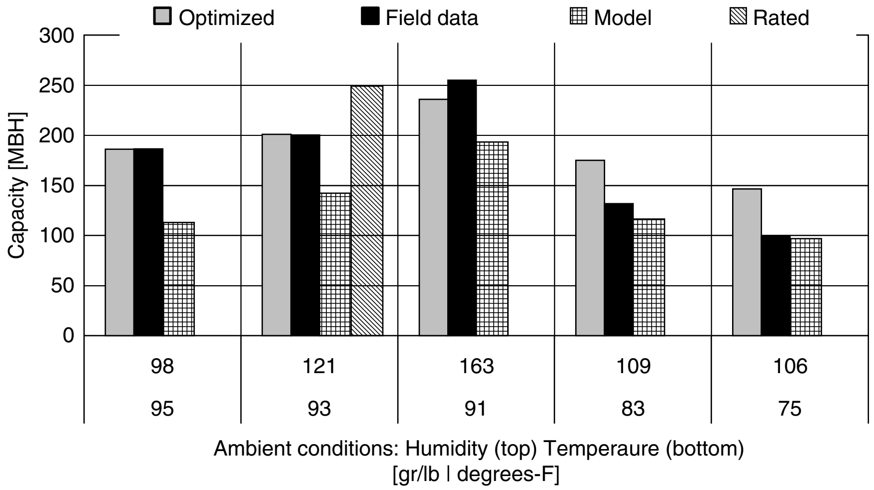


Fig. 1 Comparison of cooling capacity.

efficiency potential of desiccant equipment. Even so, it does (in our opinion) represent a "typical" commercial installation. The simple gas boiler control does not have the ability to vary heat output or gas consumption according to the need for dehumidification. The boiler is either full on or shut off, and the data clearly shows it unnecessarily operates full-on almost constantly. Our data indicates that less than 60% of the natural gas energy consumed by the unit is actually utilized. Likewise, the evaporative cooler is not nearly as effective as it should be. The analysis also indicates the possibility of moisture carry-over from the regeneration side of the evaporative cooler to the process side via the heat wheel.

Control of the unit is based simply on supply air temperature, and to a lesser degree, humidity. The data clearly shows that the control sequence does not take into account the cooling needs of the building; it aims merely to supply air at a fixed temperature regardless of whether additional mechanical cooling or reheating is necessary downstream.

COMPUTER MODELING

The most complete, representative, accurate, and reliable contiguous sets of data were used to develop, calibrate, and validate an hourly computer model. These sets included 55 days of hourly data from various periods of the project—a total of some 46,000 data points. Performance was evaluated using results from these sets of screened field data and a full-year set of computer model results as driven by the serving utility company's typical 30-year hourly weather data.

The hourly computer model consists of a set of submodels for each of the components of the system. These component submodels, such as the evaporative cooler and the desiccant wheel, are assembled together to

Dem–Dry

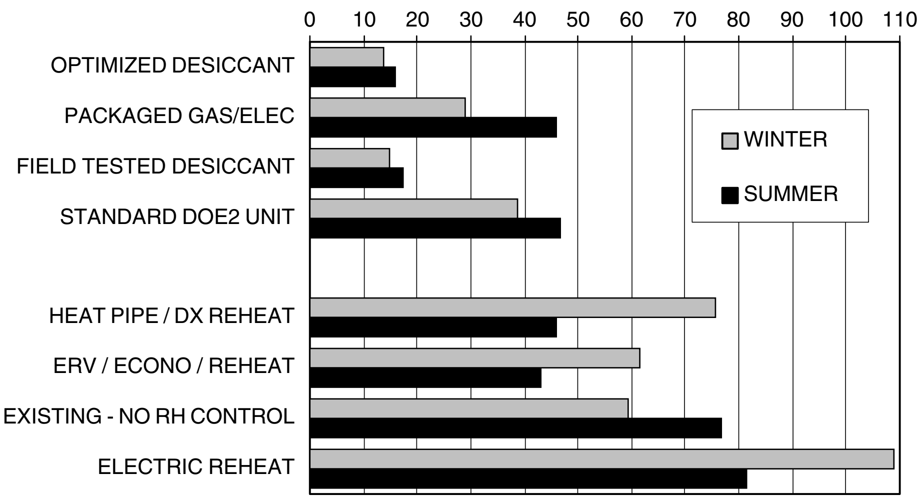


Fig. 2 Comparison of peak electric use (kW).

simulate the performance of the equipment as a whole. Each submodel was used to simulate the performance of a single component for each of the 8760 h in the typical weather year.

As a final check, the results of the model were compared against the standard DOE-2.1e hourly simulation software developed by the U.S. Department of Energy. A static comparison and validation was also performed at the outdoor temperature and humidity conditions published in the manufacturer's equipment performance specifications. Comparisons were made between the manufacturer's published data, the manufacturer's site test data taken at the time of installation, the collected field data, the computer model, and the theoretical best-case performance.

Two baseline options were developed to simulate comparable all-electric dehumidification equipment commonly used in the commercial sector. The baseline computer model simulates the existing electric-DX overcooling and electric reheating system; and alternatively, two all-electric packaged roof-top system configurations that satisfy the Florida Energy Code criteria of minimizing or avoiding the use of new energy for reheat. The first unit incorporates an energy recovery wheel (ERV) and an economizer function, and the second is equipped with wrap-around heat pipes and condenser waste heat recovery. The results of these models were also checked against the standard DOE-2.1e hourly simulation software (Fig. 2).

RESULTS

The desiccant unit as installed delivers less cooling and dehumidification capacity than the manufacturer's rating, and much less than it would if the equipment designs were optimized and the installation was commissioned. The unit consumes less energy than rated; however, it consumes considerably more than it would with optimization and commissioning. While the measured energy efficiency at peak load conditions is better than rated, the data clearly shows that this rating is not representative of long-term field performance. In contrast to all-electric cooling equipment, the efficiency of this type of unit tends to decrease as conditions become less humid and cooler. Because peak load conditions are experienced only a fraction of the time, the average efficiency is considerably lower than the rated efficiency. Furthermore, the measured decline in performance with decreasing cooling load—when dehumidification is most critical—is more severe than would be expected.

On the plus side, the primary benefit to the customer of installing the unit has been decreased humidity and increased ventilation for building occupants. The desiccant unit has provided this improvement at annual energy and maintenance savings of about 30% per year, as compared with achieving a similar improvement with the existing all-electric overcool and reheat equipment. The desiccant unit could provide the same level of comfort as existed before its installation (no improvement in humidity or ventilation) at annual energy and maintenance savings of about 16%. The peak demand of the desiccant unit is 15 kW, as compared with 77 kW for the existing equipment. The incremental cost of the desiccant installation will pay back in roughly 7 years.

Two all-electric packaged rooftop system alternatives that satisfy the Florida Energy Code criteria of minimizing or avoiding the use of new energy for reheat were also compared. Peak demand during cooling mode would be about 45 kW. A gas/electric package unit (not desiccant) would have provided an annual savings of about 30%, and a peak demand reduction from 77 to 46 kW. Any of these three options would pay back in about 5 years.

The potential savings available from optimization and field commissioning of the existing desiccant unit is an additional 25% per year, increasing the total savings to about 42% as compared with the baseline.

ACCURACY

Minor inaccuracies in the results of the computer modeling arise mostly from the assumption of linear behavior and use of linear equations in the model. Unlike most other HVAC components, the combined heat and mass transfer occurring in the desiccant wheel experiences hysteresis and nonlinear transients. For example, during relatively humid conditions, the wheel can remove significantly more humidity from the process air than it expels in regeneration. The wheel "stores" moisture in this manner typically over a period that can last hours, and sometimes days. Nonetheless, the average error between the measured field data and the computer predictions is just 7%.

Minor errors in the field data propagated from a number of sources: temperature sensor calibration error of ± 0.8 to ± 1.3°F, plus airflow measurement error of ± 50 fpm, plus dimensional measurement error of ± 0.5 in., plus relative humidity measurement error of $\pm 4\%$ rh. These field data errors result in a sensible cooling capacity error of 11%, a dehumidification capacity error of 25%, a total unit cooling capacity error of 15%, and an energy efficiency error of 18%. These errors were inherent to the sensors and equipment used, the use of "point" rather than "averaging" RTD sensors, the sometimes large differential between point sensor reading and bulk flow conditions, and the different data averaging and sampling rates of the K20 and CS data loggers.

CONCLUSIONS

1. The long-term as-installed performance of typical desiccant HVAC equipment may be less than expected in terms of both delivered capacity and energy efficiency.
2. Engineered improvements to the design and installation of typical desiccant HVAC equipment can provide large performance and cost benefits.
3. Field monitoring and computer analysis of HVAC equipment performance can reveal many cost-effective energy-saving measures.

Distributed Generation

Paul M. Sotkiewicz
Public Utility Research Center, University of Florida, Warrington College of Business, Gainesville, Florida, U.S.A.

Ing. Jesús Mario Vignolo
Instituto Ingeniería Eléctrica, Universidad de la República, Montevideo, Uruguay

Abstract
Distributed generation (DG) is generally thought of as small-scale generation that is used on site and/or connected to the distribution network. Distributed generation development has been driven by technological changes, the availability of inexpensive natural gas, the evolution of electricity competition, and perhaps most of all by the need for extremely reliable electricity supply and combined heat and power (CHP) applications. Distributed generation technologies range from small microturbines and fuel cells to larger reciprocating engines and simple cycle gas turbines. Although DG is not competitive in most applications with grid-supplied power, it has benefits such as increased thermal efficiency in co-generation applications, enhanced reliability, the potential to reduce system losses, the potential to delay or avert new infrastructure investment, and lower emissions compared with traditional coal- and oil-fired technologies.

INTRODUCTION: SOME HISTORY AND EVOLUTION TOWARD DISTRIBUTED GENERATION

Distributed generation (DG) is generally thought of as small-scale generation that is used on site and/or connected to the distribution network. Historically, the type of technologies employed has varied but generally is limited to small engines or combustion turbines fueled by diesel, gasoline, or natural gas and expensive to run relative to grid-supplied power.[1] More recently, intermittent renewable resources such as solar photovoltaic (PV), small hydro, and wind have been thought of as DG that is seen as being deployed to reduce overall emissions. Consequently, small-scale, fossil-fired generation was seen, and still is seen, as primarily providing reliable backup generation in the event of grid-supplied power interruptions, with an estimated 70% of diesel distributed generators in the United States being used for emergency purposes.[2] In contrast, the electricity industry was historically seen as possessing economies of scale in the production and delivery of power. Such economies of scale necessitated larger and larger generating facilities to meet the increasing demand. This brings us to the power system of today, where we have large, centrally dispatched power stations that are connected to one another and to consumers by the high-voltage transmission system, eventually leading to lower distribution voltages and consumers of power.

Keywords: Distributed generation; Reliability; Distribution networks; Distributed generation policies.

Several developments, however, have made the idea of DG not only possible, but potentially desirable. The first development is the technological change relating to costs and economies scale that came to fruition in the 1990s: combined cycle, natural gas technology. In Fig. 1, the change in economies of scale compared with historical trends is a fundamental shift toward smaller, lower-cost generating units.

The second development, as indicated by proponents of DG in Conseil International des Grands Résaux Électriques (CIGRE) (in English the International Council on Large Electric Systems) Working Group 37.23,[3] was the availability of relatively inexpensive natural gas supplies, which made potential DG technologies more affordable to operate. Consequently, then it would be possible, as argued in International Energy Agency (IEA)[2] and CIGRE Working Group 37.23,[3] for DG to operate at costs competitive to that of traditional central-station power while averting, deferring, or reducing network costs.

The third development, aided by the previous two developments and described by Hunt and Shuttleworth,[4] is the policy change around the world, moving from vertically integrated monopolies toward more competitive market structures in the generation sector, allowing for more diverse ownership of generating assets that would compete to drive the price of electricity down.

The last development, driven by ongoing environmental policy, is the idea that DG can help countries reduce emissions, especially carbon emissions.[5] Natural gas-fired technologies have lower carbon emissions than traditional coal-fired technologies but higher emissions than renewable technologies, which have zero carbon emissions.

The remainder of the entry is organized as follows. First, we will discuss how DG is defined and contrast that

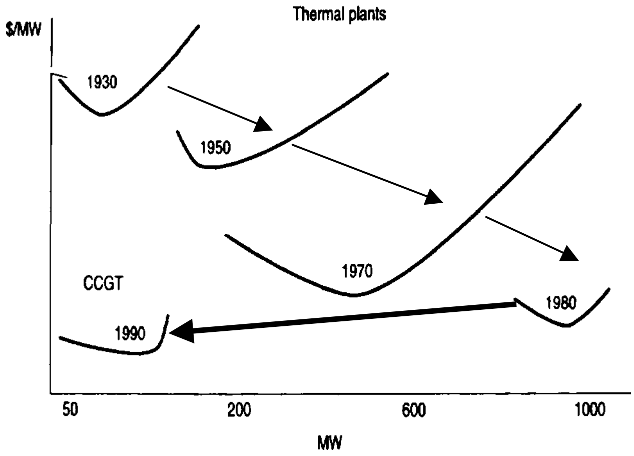


Fig. 1 Generating plants costs curves concerning power (1930–1990).
Source: From Wiley (see Ref. 4).

with other notions of distributed resources. Next, we will briefly outline the types of technologies that are deployed as DG and summarize their cost characteristics. Following that, we will discuss the potential benefits attributed to DG and provide some cautions about overstating the benefits. Finally, we will discuss policies affecting DG and provide concluding remarks.

WHAT IS DISTRIBUTED GENERATION?

Many terms have emerged to describe power that comes from sources other than from large, centrally dispatched generating units connected to a high-voltage transmission system or network. In fact, there is no clear consensus as to what constitutes DG.[2,6]

Congrès International des Réseaux Electriques de Distribution (CIRED) (in English the International Conference on Electricity Distribution) Working Group No. 4[6] created a questionnaire that sought to identify the current state of dispersed generation in the various CIRED member countries. Response showed no agreement on a definition, with some countries using a voltage level definition while others considered direct connection to consumer loads. Other definitions relied on the type of prime mover (e.g., renewable or co-generation), while others were based on non-centrally dispatched generation.

This diversity is also reflected in the CIGRE Working Group 37.23[3] definition, which characterizes dispersed generation as not centrally planned or dispatched, connected to lower-voltage distribution networks, and less than 50–100 MW.

The World Alliance for Decentralized Energy (WADE)[7] defines decentralized energy (DE) as:

- Electricity production at or near the point of use, irrespective of size
- Technology or fuel used, both off grid and on grid, including (1) high-efficiency cogeneration on any scale, (2) on-site renewable energy, and (3) energy recycling systems powered by waste gases, waste heat, and pressure drops to generate electricity and/or useful thermal energy on site.

The IEA[2] defines DG as the following:

Distributed generation is a generating plant serving a customer on site or providing support to a distribution network, connected to the grid at distribution level voltages. The technologies include engines, small (and micro) turbines, fuel cells, and PV systems.

The IEA definition excludes wind power, arguing that it is mostly produced on wind farms usually connected to transmission, rather than for on-site power requirements. In addition to providing a definition for DG, the IEA[2] has provided nomenclature for other dispersed, distributed, or DE resources that we outline below for completeness and to alert the reader of the different terms that are often used with respect to DG. It should be noted that in each of the bulleted definitions below, DG is a subset of the defined category:

- Dispersed generation includes DG plus wind power and other generation, either connected to a distribution network or completely independent of the grid.
- Distributed power includes DG plus energy storage technologies such as flywheels, large regenerative fuel cells, or compressed air storage.
- Distributed energy resources include DG plus demand-side measures.
- Decentralized power refers to a system of distributed energy resources connected to a distribution network.

For the purpose of this work, we will consider DG as generation used on site (and possibly unconnected to the distribution network) and/or connected to the lower-voltage distribution network irrespective of size, technology, or fuel used. This nomenclature encompasses the definitions of IEA[2] and WADE.[7]

DG Technologies

Reciprocating Engines

Reciprocating engines, according to IEA,[2] are the most common form of DG. This is a mature technology that can be fueled by either diesel or natural gas, though the majority of applications are diesel fired. The technology is capable of thermal efficiencies of just over 40% for electricity generation, relatively low capital costs, but relatively high running costs, as shown in Table 1. The technology is also suitable for backup generation, as it can be started quickly and without the need for grid-supplied power. When fueled by diesel, this technology has the

Table 1 Cost and thermal efficiencies of distributed generation (DG) technologies, inclusive of grid connection costs and without combined heat and power (CHP) capability

	Installed cost ($/kW)	O&M (c/kWh)	Efficiency (%)	Levelized cost (c/kWh)
Simple cycle gas turbine	650–900	0.3–0.8	21–40	6–9
Microturbines	1000–1300	0.5–1.0	25–30	7–9
Diesel engines	350–500	0.5–1.0	36–43	7–11
Gas engines	600–1000	0.7–1.5	28–42	6–9
Fuel cells	1900–3500	0.5–1.0	37–42	11–14
Solar photovoltaic (PV)	5000–7000	0.1–0.4	NA	34.5–46.0
Small hydro	1450–5600	0.7	NA	3.5–8
Wind	790	NA	NA	7.6

Source: From OECD/IEA (see Ref. 2) except for the wind which is from AWEA (see Ref. 8) and Small Hydro from WADE (see Ref. 7). Levelized cost numbers assume 60% capacity factor except for Solar PV from WADE (see Ref. 7) at 1850 h/year, Small Hydro from WADE (see Ref. 7) at 8000 h/year, and wind at 39% capacity factor.

highest nitrogen oxide (NO_x) and carbon dioxide (CO_2) emissions of any of the DG technologies considered in this entry, as shown in Table 2.

Simple Cycle Gas Turbines

This technology is also mature, deriving from the use of turbines as jet engines. The electric utility industry uses simple cycle gas turbines as units to serve peak load, and they generally tend to be larger. Simple cycle gas turbines have the same operating characteristics as reciprocating engines in terms of startup and the ability to start independently of grid-supplied power, making them suitable as well for backup power needs. This technology is often run in combination for combined heat and power (CHP) applications, which can increase overall thermal efficiency. Capital costs are on par with those of natural gas engines, as shown in Table 1, with a similar operating and levelized cost profile. The technology tends to be cleaner, as it is designed to run on natural gas, as shown in Table 2.

Microturbines

This technology takes simple cycle gas technology and scales it down to capacities of 50–100 kW. The installed costs are greater than for gas turbines, and the efficiencies are lower as well, as shown in Table 1. Microturbines are much quieter than gas turbines, however, and have a much lower emissions profile than gas turbines, as shown in Table 2. The possibility also exists for microturbines to be used in CHP applications to improve overall thermal efficiencies.

Fuel Cells

Fuel cell technology is also fairly new and can run at electrical efficiencies comparable to those of other mature technologies. Fuels cells have the highest capital cost

Table 2 Emission profiles of distributed generation (DG) technologies

Technology	lbs. NO_x/MWh	lbs. NO_x/mmBtu	lbs. CO_2/MWh	lbs. CO_2/mmBtu
Average coal boiler 1998	5.6	0.54	2115	205
Combined cycle gas turbine 500 MW	0.06	0.009	776	117
Simple cycle gas turbine	0.32–1.15	0.032–0.09	1154–1494	117
Microturbines	0.44	0.032	1596	117
Diesel engines	21.8	2.43	1432	159
Gas engines	2.2	0.23	1108	117
Fuel cells	0.01–0.03	0.0012–0.0036	950–1078	117
Solar photovoltaic (PV)	0	0	0	0
Small hydro	0	0	0	0
Wind	0	0	0	0

Source: From RAP (see Ref. 9).

Dem–Dry

among fossil-fired technologies and consequently have the highest levelized costs, as shown in Table 1. Offsetting that, the emission footprint of fuel cells is much lower than that of the other technologies, as shown in Table 2.

Renewable Technologies

We discuss three major types of renewable energy technologies here: solar PV, small hydro, and wind. Each of these technologies is intermittent, in that it is dependent upon the sun, river flows, or wind. Consequently, these technologies are not suitable for backup power, but also have no fuel costs and have a zero emissions profile, as shown in Table 2. The capital costs vary significantly among the technologies, however, and operating conditions over the year affect their respective levelized costs. Solar PV is by far the most expensive in both capital costs and levelized costs, as shown in Table 1. Capital costs for small wind are much lower, but levelized costs are in the range of more traditional technologies, as shown in Table 1. Small hydro capital costs can vary widely, with levelized costs reflecting the same variation.

The Role of Natural Gas and Petroleum Prices in Cost Estimates

The levelized cost figures in Table 1 make assumptions about the price of natural gas and diesel. As shown in U.S. Energy Information Administration (USEIA),[10] the prices of natural gas and petroleum products have risen substantially in recent years relative to the time the levelized cost estimates have been calculated. Consequently, if the forecasts in USEIA[10] turn out to be relatively accurate, the levelized cost of all the fossil technologies will be greater than the revelized costs stated here, all else equal.

Potential Benefits of Distributed Generation

Distributed generation has many potential benefits. One of the potential benefits is to operate DG in conjunction with CHP applications, which improves overall thermal efficiency. On a stand-alone electricity basis, DG is most often used as backup power for reliability purposes but can also defer investment in the transmission and distribution network, avert network charges, reduce line losses, defer the construction of large generation facilities, displace more expensive grid-supplied power, provide additional sources of supply in markets, and provide environmental benefits.[11] Although these are all potential benefits, however, one must be cautious not to overstate the benefits, as we will discuss as well.

Combined Heat and Power (CHP) Applications

Combined heat and power, also called cogeneration, is the simultaneous production of electrical power and useful heat for industrial processes, as defined by Jenkins et al.[12] The heat generated is used for industrial processes and/or for space heating inside the host premises or is transported to the local area for district heating. Thermal efficiencies of centrally dispatched, large generation facilities are no greater than 50% on average over a year, and these are natural gas combined cycle facilities.[9] By contrast, cogeneration plants, by recycling normally wasted heat, can achieve overall thermal efficiencies in excess of 80%.[7] Applications of CHP range from small plants installed in buildings (hotels, hospitals, etc.) up to big plants on chemical works and oil refineries, although in industrialized countries, the vast majority of CHP is large industrial CHP connected to the high-voltage transmission system.[2] According to CIGRE Working Group 37.23,[3] the use of CHP applications is one of the reasons for increased DG deployment.

Table 3 shows the costs of DG with CHP applications and their levelized costs. Compared with the levelized costs of stand-alone electricity applications, these costs are lower, especially at high capacity factors (8000 h), showing evidence of lower costs along with greater efficiency in spite of the higher capital cost requirements.

Impact of DG on Reliability (Security of Supply)

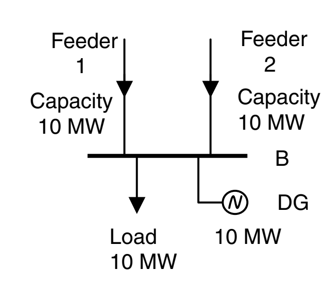
It seems quite clear that the presence of DG tends to increase the level of system security. To confirm this idea, the following example is presented.

Table 3 Distributed generation (DG) technology costs, inclusive of combined heat and power (CHP) infrastructure

			Levelized cost (c/kWh)	
	Installed capital cost ($/kW)	**O&M (c/kWh)**	**8000 h/year**	**4000 h/year**
Simple cycle gas turbine	800–1800	0.3–1.0	4.0–5.5	5.5–8.5
Combined cycle gas turbine	800–1200	0.3–1.0	4.0–4.5	5.5–6.5
Microturbines	1300–2500	0.5–1.6	5.0–7.0	7.0–11.0
Reciprocating engines	900–1500	0.5–2.0	4.5–5.5	6.0–8.0
Fuel cells	3500–5000	0.5–5.0	9.0–11.5	14.5–19.5

Sources: From OECD/IEA (see Ref. 2) and WADE (see Ref. 7).

Dem–Dry



Equipment	For (forced outage rate)
Feeder 1	0.04
Feeder 2	0.04
DG	0.20

Fig. 2 Security of supply example with distributed generation (DG).

Fig. 2 shows a very simple distribution network. It consists of two radial feeders, each with 10 MW of capacity, that feed busbar B. A constant load of 10 MW is connected to B. The Forced outage rate (FOR) of the two feeders is given in the table in Fig. 2. Additionally, consider a 10 MW DG source with an availability factor of 80%.

To begin with, let us consider only the two feeders and assume that there is no distributed resource connected to busbar B. The loss of load probability (LOLP)—the probability that load is not served—is simply the probability that both feeders will be out of service at the same time, which can be calculated by multiplying the two probabilities of failure. Consequently, LOLP $=(0.04\times0.04)=0.0016$. The expected number of days in which the load experiences troubles can also be calculated by multiplying the LOLP by 365, which results in 0.584 days/year. This number can be expressed in hour/year by multiplying by 24, resulting in 14 h/year.

Now let us consider including the DG source. It has an outage rate greater than the two feeders at 0.20, but it also adds a triple redundancy to the system. Thus, we would expect the addition of the DG source to decrease the LOLP. The new LOLP is the probability that both feeders will fail and that the DG source is not available. Therefore, LOLP $=(0.04\times0.04\times0.20)=0.00032$—that is, the probability of being unable to serve load is 5 times less than before. This translates to an expected number of hours per year unable to serve load at just less than 3 h/year in our example.

Impact of Distributed Generation on Network Losses, Usage, and Investment

The presence of DG in the network alters the power flows (usage patterns) and, thus, the amount of losses. Depending on the location and demand profile in the distribution network where DG is connected and DG operation, losses can either decrease or increase in the network. A simple example derived from Mutale et al.[13] can easily show these concepts.

Fig. 3 shows a simple distribution network consisting of a radial feeder that has 2 loads (D1 and D2 at points A and B, respectively) and a generator (G) embedded at point C. The power demanded by the loads is supposed to be constant and equal to 200 kW. The power delivered by the G is 400 kW. The distance between A and B is the same as the distance between B and C. In addition, the distance between T and A is twice the distance between A and B. Moreover, we assume that the capacity of each of the sections is equal to 1000 kW. Impedances for sections AB and BC are assumed to be equal, as are the distances. The impedance on TA is assumed to be twice that of AB and BC, as the distance is double. We also assume that voltages are constant and that losses have a negligible effect on flows.

From this hypothesis, it is easy to demonstrate that the line losses (l) can be calculated by multiplying the value of line resistance (proxy for impedance) (r) by the square of the active power flow (p) through the line: $l=rp^2$.

If distributed G is not present in the network (disconnected in Fig. 4), the loads must be served from point T, with the resulting power flows, assuming no losses for the ease of illustration, of Fig. 4.

Losses in the network are $l = 4^2(2 \times 0.001)+ 2^2 \times 0.001 = 0.036$ p.u., or 3.6 kW. Additionally, the usage of the network is such that the section TA is used to 40% of its capacity (400 kW/1000 kW) and section AB is used to 20% of its capacity (200 kW/1000 kW).

Now assume that DG G is connected at point C, as shown in Fig. 5.

The resulting power flows, assuming no losses again for ease of illustration, are the following:

The losses are $l = 0.001[2^2 + 4^2] = 0.02$ p.u., or 2 kW, which is a 44% reduction in losses in the case without DG. The reduction from losses comes from transferring flows from the longer-circuit TA to the shorter-circuit BC. Moreover, because less power must travel over the transmission network to serve the loads D1

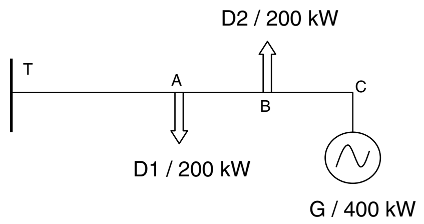


Fig. 3 A simple distribution network.

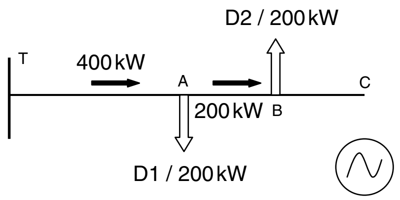


Fig. 4 Power flows without distributed generation (DG).

and D2, losses on the transmission system are reduced, all else being equal.

Additionally, the pattern of usage has changed. The usage on AB is still 200 kW, but the flow is in the opposite direction from the situation without DG. The flow on TA has been reduced from 400 to 0 kW. In effect, the DG source at C has created an additional 400 kW of capacity on TA to serve growing loads at A and B. Suppose that the loads D1 and D2 increased to 700 kW each. Without DG, this would require extra distribution capacity to be added over TA, but with DG, no additional distribution capacity is needed to serve the increased load. In short, DG has the ability to defer investments in the network if it is sited in the right location.

Finally, depending on the distribution and transmission tariff design, DG can avert paying for network system costs. This is especially true in tariff designs, in which all network costs are recovered through kWh charges rather than as fixed demand charges. This is another reason, according to CIGRE Working Group 37.23,[3] for DG deployment.

It is important to emphasize that the potential benefits from DG are contingent upon patterns of generation and end use. For different generation and end-use patterns, losses and usage would be different. In fact, losses may increase in the distribution network as a result of DG. Let G produce 600 kW, for example. In this case, losses are 6 kW—greater than the 3.6 kW losses without DG. Moreover, although DG effectively created additional distribution capacity in one part of the network, it also increased usage in other parts of the network over circuit BC. Consequently, one must be cautious when evaluating the potential for DG to reduce losses and circuit usage.

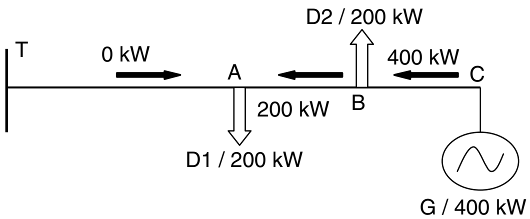


Fig. 5 Power flows and usage with generator (G) producing 400 kW.

Potential to Postpone Generation Investment

In addition to the potential network benefits and reliability (security of supply benefits), DG may bring other benefits to power systems. The first is the ability to add generating capacity in a modular fashion, which does not require building large power plants that will have excess capacity for some time and that, because of size, may be easier to site and permit and faster to complete. In this vein, Hadley et al.[14] modeled DG in the Pennsylvania-New Jersey-Maryland Interconnection (PJM) market and found the potential to displace some existing units, as well as postpone new combined cycle gas units. One must be cautious with this potential benefit, however, as the overall costs of DG may be more than central-station power.

Potential Electricity Market Benefits

In an electricity market environment, DG can offer additional supply options to capacity markets and the ancillary-services market, thereby leading to lower costs and more competition.[15] In the same vein, the owner of DG has a physical hedge against price spikes in electricity markets, which not only benefits the owner of DG but also should help dampen the volatility in the market.[2]

Potential Environmental Benefits

Finally, DG resources may have lower emissions than traditional fossil-fired power plants for the same level of generation, as shown in Table 2, depending on technology and fuel source. This is true for renewable DG technologies, of course. The benefits are potentially large in systems in which coal dominates electricity generation, as shown in Table 2. Hadley et al.[14] models DG in the PJM market and finds that DG displacing generation on the system led to lower emission levels. These reasons were cited in CIGRE Working Group 37.23[3] as determining factors for some DG deployment. Moreover, because losses may also be reduced, DG may reduce emissions from traditional generation sources as well. Additionally, customer demand for renewable energy may be driving renewable energy deployment.[16]

POLICIES AND CONCLUDING REMARKS

Distributed generation as defined in IEA[2] and WADE[7] can provide many benefits, though it is not yet quite competitive with grid-supplied power on its own. Current policies to induce DG additions to the system generally consist of tax credits and favorable pricing for DG-provided energy and services that are subsidized by government.[2] Although such policies may be effective to capture some potential benefits from DG, such as environmental benefits, they do not address the network or market benefits of DG. Only recently has serious consideration been given to considering locational pricing of network services as a way to provide better incentives without subsidies,[17,18] as recommended by IEA.[2] Moreover, only recently has DG been recognized as a potential player in wholesale power markets to provide marketwide benefits.[15] Finally, any barriers that prevent the efficient entry of DG should be reconsidered.[1,2]

REFERENCES

1. U.S. Congressional Budget Office (USCBO). *Prospects for Distributed Electricity Generation*; Congressional Budget Office: Washington, DC, 2003.
2. International Energy Agency (IEA). *Distributed Generation in Liberalised Electricity Markets*; OECD/IEA: Paris, 2002.
3. CIGRE Working Group 37.23. *Impact of Increasing Contribution of Dispersed Generation on the Power System*; Paris, France, 1999.
4. Hunt, S.; Shuttleworth, G. *Competition and Choice in Electricity*; Wiley: London, 1996.
5. World Alliance for Decentralized Energy (WADE). *World Survey of Decentralized Energy 2005*. Available at: www.localpower.org/documents_pub/report_worldsurvey05.pdf (accessed on November 2006).
6. CIRED Working Group No. 4 on Dispersed Generation (CIRED). *Preliminary Report Discussion at CIRED 1999*; Nice: France, June 2, 1999.
7. World Alliance for Decentralized Energy (WADE). *WADE Guide to DE Technologies: Cogeneration Technologies*, 2003; Available at: www.localpower.org/documents_pub/report_de_technologies.pdf (accessed on November 2006).
8. American Wind Energy Association (AWEA). *Economics of Wind Energy*, Washington, DC, AWEA Available at: www.awea.org/pubs/factsheets/EconomicsofWind-March2002.pdf (accessed on November 2006).
9. Regulatory Assistance Project (RAP). *Expected Emissions Output from Various Distributed Energy Technologies*. Available at: www.raponline.org/ProjDocs/DREmsRul/Collfile/DGEmissionsMay2001.pdf (accessed on November 2006).
10. U.S. Energy Information Administration (USEIA). Annual Energy Outlook 2006 with Projections to 2030. Available at: www.eia.doe.gov/oiaf/aeo/pdf/0383(2006).pdf (accessed on November 2006).
11. Ianucci, J.J.; Eyer, J.M.; Pupp, R.L. *DER Benefits Analysis Study: Final Report*, NREL/SR-620-34636; National Renewable Energy Laboratory: Golden, CO, September 2003.
12. Jenkins, N.; Allan, R.; Crossly, P.; Kirschen, D.; Strbac, G. Embedded Generation. In *IEE Power and Energy Series 31*; IEE: London, 2000.
13. Mutale, J.; Strbac, G.; Curcic, S.; Jenkins, N. Allocation of losses in distribution systems with embedded generation. IEE Proc.-Gener. Transm. Distrib. **2000**, *147* (1), 1–8.
14. Hadley, S.W.; Van Dyke, J.W.; Stovall, T.K. *The Effect of Distributed Energy Resource Competition with Central Generation*, ORNL/TM-2003/236; Oak Ridge National Laboratory, U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy: Oak Ridge, TN, October 2003.
15. Sotkiewicz, P.M. *Nodal Pricing and MW-Mile Methods for Distribution: Have We Uncovered Markets or Elements for Wholesale Markets*, Presented at: Harvard Electricity Policy Group 42nd Plenary Session, La Jolla, CA, March 2, 2006; Available at: www.ksg.harvard.edu/hepg.
16. Hyde, D. Will consumers choose renewable energy, and should you care one way or the other?. *IEEE Power Engineering Review*; IEEE: London, May 1998.
17. Sotkiewicz, P.M.; Vignolo, J.M. Nodal pricing for distribution networks: efficient pricing for efficiency enhancing DG. *IEEE Transactions on Power Systems*; IEEE: London, 2006; Vol. 21, No. 2, 1013–1014.
18. Sotkiewicz, P.M.; Vignolo, J.M. Allocation of fixed costs in distribution networks with distributed generation. *IEEE Transactions on Power Systems*; IEEE: London, 2006; Vol. 21, No. 2, 639–652.

Dem–Dry

Distributed Generation: Combined Heat and Power☆

Barney L. Capehart
Department of Industrial and Systems Engineering, University of Florida College of Engineering, Gainesville, Florida, U.S.A.

D. Paul Mehta
Department of Mechanical Engineering, Bradley University, Peoria, Illinois, U.S.A.

Wayne C. Turner
Industrial Engineering and Management, Oklahoma State University, Stillwater, Oklahoma, U.S.A.

Abstract

Distributed generation (DG) is electric or shaft power generation at or near the site of use as opposed to central power station generation. Combined heat and power (CHP) takes advantage of this site location to recover the normally wasted thermal energy from power generation and utilizes it beneficially to increase the total system efficiency. This article explores the rapidly developing world of DG and associated CHP. First the article shows why DG is necessary in the U.S. power future and that DG is going to happen. Then, the article briefly looks at the different technologies that might be employed and their relative advantages and disadvantages. The article then explores who should be the major designers and implementers of DG and CHP technologies, and develops a strong argument that in many cases this should be an Energy Service Company (ESCO). Finally, the reasons for selecting either an independent ESCO or a local utility-affiliated ESCO are discussed, and in particular, opportunities for the local utility ESCO (the local grid) to be a major moving force in this effort are examined in depth.

INTRODUCTION

Distributed generation (DG) is electric or shaft power generation at or near the user's facility as opposed to the normal mode of centralized power generation and utilization of large transmission and distribution lines. Since DG is at or near the user's site, combined heat and power (CHP) becomes not only possible, but advantageous for many facilities. The CHP is the simultaneous production of electric or shaft power, and the utilization of the thermal energy that is "left over" and normally wasted at the central station generating site. Since DG means the power is generated at the user's site, CHP can be used to beneficially recover "waste" heat, and provide the facility with hot water, hot air, or steam, and also cooling through the use of absorption chillers.

Normal power generation using a steam Rankine cycle (steam turbine) is around 35% efficient for electric power production and delivery to the using site. The DG with its associated CHP potential means, the total system efficiency can be improved dramatically and sometimes even doubled. Thus, even though DG cannot usually beat the electrical generation efficiency of 35–40% (at the central station), it can save the user substantial amounts of money through recovery of the thermal energy for beneficial use at the site. In addition, there are many other potential benefits from DG/CHP, discussed below.

Thus, the user can choose the objective he/she desires and will likely find a technology in this list that meets that objective. Objectives might include power production that is environmentally friendly, cost effective, more reliable, or yields better power quality. Each of these candidate technologies will be briefly explored in the sections that follow.

Why Distributed Generation?

To explore why DG should be (and is becoming) more popular, the question of why DG needs to be addressed from the perspective of the user, the utility, and society in general. Each of these perspectives is examined below.

The user might desire more reliable power. This can occur with DG, especially, when it is connected to and backed-up by the grid. The user might desire better quality power which can result, because there will be fewer momentary interruptions and possibly better voltage consistency. Often, the user desires better economics (cheaper power), which is quite possible with DG when CHP is employed. Finally, there could be a competitive

☆ This entry originally appeared as "Distributed Generation and Your Energy Future" in *Cogeneration and Distributed Generation Journal*, Vol. 18, No. 4, Fall 2003. Reprinted with permission from AEE/Fairmont Press.

Keywords: Distributed generation (DG); Combined heat and power (CHP) systems; Cogeneration; Self-generation; On-site generation; Utility-affiliated ESCOs; Reciprocating engines; Gas turbines; Fuel cells; Photovoltaic cells; Wind turbines.

advantage during a utility power outage when the user has power and the competition does not.

The utility might desire less grid congestion and less future grid construction, both of which DG definitely yields. The utility may be able to hold on to a customer better, if the customer has DG and CHP at their site. Certainly, this is true if the utility constructs and runs the local power facility. Today's technology is capable of allowing the utility to remotely dispatch literally hundreds of DG units scattered in the grid, which could dramatically improve their ability to handle peak load and grid congestion problems.

Society in general likes DG/CHP, which can provide strong environmental advantages (when the appropriate technology is utilized). For example, CHP means less total fuel will be consumed placing less strain on the environment. Also, DG means less grid construction, and "green technology" (wind, photovoltaics, etc.) can be used if desired. Since less total fuel is consumed, there will be reduced reliance on imported oil and gas, and an improved balance of payments for the United States.

Basic Philosophies

In the development to follow, some basic objectives are assumed. They are:

- Any expenditure of funds should be cost effective
- Any change should be good for all parties, or at least be of minimal harm to any party
- Existing partnerships that are working should be maintained if at all possible.

These assumptions should be clear as stated above; but more explanation will follow as the arguments are developed.

Existing and Future Markets for DG

Resource dynamics corporation has estimated that the installed base of DG units greater than 300 kW in size is at least 34 GW (34×10^9 W). The Gas Research Institute (GRI) estimates that the installed DG capacity in the United States of all sizes is 75 GW. Just over 90% of the installed units are under 2 MW (2×10^6 W) and well over 90% of the installed units are reciprocating engines. The use of DG power plants for back-up purposes is growing steadily at 7% per year while other DG applications for baseload and peaking requirements are growing at 11 and 17%, respectively. Resource Dynamics Corporation says there is the potential to double the installed DG capacity by adding as much as 72 GW by 2010. All these sources confirm that, there is already a large base of DG units nationally, and that the growth will be significant. Distributed generation is happening and will continue to happen.

Without trying to stratify it with exact numbers, the potential market could be broken into three components. They are (1) large and medium, (2) small, and (3) smaller. Each of these categories is examined briefly below.

The large and medium market is often 25 MW and larger (sometimes hundreds of megawatts) and is a mature market because there have been plants operating for many years. Typically these are in the larger process industries such as petroleum refining, pulp and paper, and chemical plants. Steam production may be in the range of hundreds of thousands of pounds per hour. While there are many such operating plants today, this mature market probably still offers the largest immediate growth potential. There is much more that could be done.

The small market will range somewhere between 50 (or 100) kW and 25 MW. These might be plants that need significant steam and could easily add a topping or bottoming steam turbine to become DG–CHP. Important to their success is the need for thermal energy and electricity (or shaft power) and the relative sizes of those needs might dictate which technology is appropriate. This market is virtually untouched today and the management/maintenance talents in these facilities might easily support this DG/CHP technological addition to their needs. Some facilities of this size will not have the management backing and maintenance talent that it takes to make these DG/CHP systems operate successfully. Those facilities would likely seek "outsourcing" for the power plant. The growth possibility here is extremely large, but will likely take a few years to realize its full potential.

The smaller market would include those small manufacturing plants or commercial facilities that need less than 50 (or 100) kW and do not have large thermal needs. These plants and facilities likely do not possess the management backing and desire or maintenance talent it takes to run them. The market potential here is tremendous in numbers of applications, but small in numbers of total megawatts. Finally, there is a significant drop in economies of scale somewhere around 200–500 kW, so the economics here would not be as exciting as the other two markets. Thus, the authors' opinion is that this market will not be as robust in the near to immediate future. Note that this could change overnight, if a local Energy service company (ESCO) such as the utility ESCO offered to design, install, run, and dispatch these units.

DG Technologies

There is a wide range of technologies possible for DG. They include:

- Reciprocating engines—diesel, natural gas, and dual fuel
- Gas turbines—microturbines, miniturbines, and large turbines
- Steam turbines

Table 1 Overview of distributed generation (DG) technology

Technology	Pros	Cons
Fuel cell	Very low emissions	High initial investment
	Exempt from air permitting in some areas	Only one manufacturer producing commercially available units today
	Comes in a complete "ready to connect" package	
Gas turbine	Excellent service contracts	Requires air permit
	Steam generation capabilities	The size and shape of the generator package is relatively large
	Mature technology	
Micro turbine	Low initial investment	Relatively new technology
	High redundancy with small units	Requires an air permit
	Low maintenance cost	Possible synchronization problems at large installations
	Relatively small size	
	Installation flexibility	
Engine	Low initial investment	High maintenance cost
	Mature technology	Low redundancy in large sizes
	Relatively small size	Needs air permit
Photovoltaics	Low operations and maintenance (O&M) costs	Very expensive initially
	Environmentally friendly	Very large footprint
		Sun must shine
		Battery storage usually needed
Wind	Low to medium O&M costs	Large footprint
	Environmentally friendly	Wind must blow

- Fuel cells
- Photovoltaic cells
- Wind turbines
- Storage devices (batteries or flywheels).

The following table briefly summarizes the pros and cons of these different DG/CHP technologies (Table 1). If more detailed information is needed, the authors recommend Capehart, et al[1]; Turner[2]; or Petchers[3].

The table above demonstrates that there is a wide range of technologies available. Some are environmentally friendly, some are not. Some are more economically feasible, while others are extremely expensive. Some use mature technologies, while others are still somewhat of a gamble. What is badly needed for this market to mature are more ESCOs that are broadly experienced in DG/CHP applications and that are good at all of the above technologies, including the nontraditional approaches. Their tool sack contains all of these technologies and they know when and how to apply each of these technologies. To our knowledge, today only a few ESCOs can claim this broad a talent base.

Who?

Thus far, we have demonstrated that there is a significant market projected for DG/CHP systems and that this market needs to be satisfied. We have also shown there is a wide range of technologies that is available. What is needed is someone to "make this happen." Rather obviously, there are about three groups that could make this happen. They are

The users themselves
Energy service companies (ESCOs)
— Independent ESCOs (consultants)
— Utility affiliated ESCOs.

This section examines each of these groups, and shows how they might contribute to the expanded need for DG/CHP. One consideration in evaluating the potential success of a DG/CHP project is the goal alignment of the participants, where the goals of the user or the facility are compared to the goals of the organization that is implementing the project. The closer these goals match up, or align, the more likely the DG/CHP project is to succeed.

The Group of Users Themselves

The users' goals are to have a DG/CHP project that provides an appropriate solution to their needs for electric or shaft power, and probably heat; works well for them both in the short term and the long term; and maximizes their economic benefit from this investment.

The user knows its process better than anyone else. This is a real advantage of doing DG/CHP projects in house and leads to the best economics if it works. Finally, the goal alignment for this group is the best of the three groups, as it is the user itself doing the job.

For this to work, the user must have a staff of technically qualified people who can analyze potential technologies, evaluate the options, select the best technology for their application, permit and install the equipment, and operate the DG/CHP system in a manner which produces the desired results. In addition, management and maintenance must both commit to the project. This often will not occur, if they wish to devote their time and efforts to building better products, delivering better services, or expanding into new products or services. Another disadvantage is that a very large capital investment is normally required and many plants and facilities simply do not have the necessary capital. Finally, these projects would involve grid interconnections and environmental permitting. Many plants and facilities are very unfamiliar with these requirements.

However, if the facility or plant does have a committed and skilled management and staff that can select, permit, install, finance and operate the DG/CHP system, this approach will most likely provide them with the highest rate of return for this kind of project.

The ESCO Group

For facilities that cannot or do not want to initiate and implement their own DG/CHP projects, the involvement of an ESCO, is probably their most appropriate alternative. Energy Service Companies bring a very interesting set of talents to DG/CHP projects. The right ESCO knows how to connect to the grid, what permits are required, and how to obtain those permits. The right ESCO knows all of the technologies available and how to choose the best type and size to utilize for this application. The right ESCO is a financial expert that knows all of the financing options available and which might be the best. Often, this means they have partnered with a financing source and have the money available with payback based on some mutually agreeable terms (interest bearing loan, shared savings, capital lease, true lease, etc.).

One of the disadvantages of using an ESCO is, they are sometimes "in a hurry." When this happens, the project design is not as well done as it should be, and they may leave before the equipment is running properly. Commissioning becomes extremely important here. Another disadvantage is that some ESCOs choose the same technology (cookie cutting) for all projects. A certain type turbine made by a particular company is always chosen, when this may not always be the best solution. If the ESCO approach is to work, all technologies must be considered and the best one chosen. For this group, goal alignment is not the best as the user no longer is in charge and the relationship is likely to be of limited duration (outsourcing being a possible exception).

However, if an "ideal" ESCO can be found and utilized, this approach offers a very satisfactory arrangement.

Independent ESCOs (Consultants)

The goals of the independent ESCO are typically to sell the customer a technology solution that the ESCO is familiar with; get the equipment installed and checked out quickly; maximize their profit on the project; and in the absence of a long term contract to provide maintenance or operating assistance to get out as quickly as they can. Sometimes the independent ESCOs goals do not line up that closely with the customer's. The independent ESCO may try to sell the customer a particular piece of equipment that they are most familiar with, and may be the one that gives them the largest profit. If there is not going to be any long-term contract for the ESCO, then they want to get the project completed as quickly as possible, and then get out. This may leave the user with a DG/CHP system that is not thoroughly checked out and tested, and leaves the user to figure out how to operate the system and how to maintain it. The project ESCO team may then depart the facility, and return to their distant office, which may be in a very different part of the country.

However, as long as the user is willing to pay the ESCO for continuing their support, the ESCO is almost always willing to do that. Unless the user is willing to pay for a part time or full time person to remain at their facility, they will have to deal with the ESCO by phone, FAX, FedEx, or Email.

One of the other potential problems with an independent ESCO is the question of its permanence. Will it be around for the long term? Historically, making the comparison of current DG/CHP ESCOs to the solar water heating companies of the 1970s and 1980s, leads to the concern that some of the DG/CHP ESCOs may not be around for the long term. Very few of the companies that manufactured, sold, or installed solar water heating systems at that time are still around today. Many of these solar water heating companies were gone within a few years of the customers purchasing the systems. Most of these companies were actually gone long before the useful lives of the solar systems had been reached. Repair services, parts, and operating advice were often no longer available, so many solar water heating system users simply stopped using them, or removed the systems. Based on this history, selection of an ESCO that is likely to be around for the long term is an important consideration.

Utility-Based ESCOs

Next, consider a utility-based ESCO. Utility-affiliated ESCOs have goals similar to the independent ESCOs' goals in many respects, but the big difference is that the utility is a permanent organization that is local is there for the long term, and is interested in seeing the user succeed, so that they will be an even better customer in the long run. Also, since the utility and the affiliated ESCO are local, they can send someone out periodically to check on the facility and the DG/CHP project to help answer questions and make sure the project is continuing to operate successfully. The utility is financially secure, stable, and, in most instances, is regarded by the community as an honest and trustworthy institution.

This ESCO now is an independent branch of the local utility. If they have the full set of tools (knowledge of all the technologies) then their advantages include all those listed above. In addition, there is much better goal alignment. They will be there as a partner as long as the wires are connected and that likely is almost forever. Thus, both parties want this to work. They are the grid, so the grid interconnection is not as much of a problem. The user and the utility have been partners for years. This would change the relationship; not destroy it. (The devil you know vs the one you do not.) Finally, this is what they do (almost).

One limitation of the utility-based ESCO is that they must change their mindset of "sell as much electricity as possible," and recognize that there is a lot of business and income to be captured from becoming an energy service organization. Someone is going to do these DG/CHP projects; the utility revenue base is most enhanced when they do it and when the project is successful. Their services could involve design, installation, start up, commissioning, and passing of the baton to the user or they might run it themselves (outsourcing).

Now, if the local utility company ESCO can take advantage of the opportunities they have, then they have a lot to offer to facilities and plants that are interested in working with them to put in DG and CHP systems. The old Pogo adage "We are surrounded by insurmountable opportunities" is always around in these situations. Another old saying would describe this DG/CHP opportunity for utility-affiliated ESCOs as "the business that is there for them to lose." The utility-affiliated ESCOs need to aggressively pursue these opportunities.

Some Local Utility ESCO Successes

A very good example of a local utility ESCO success story comes from the experiences shared by AmerenCILCO, a utility company in central Illinois, This company has experiences with both DG and CHP projects.

DG Only Projects

AmerenCILCO has extensive experience in using reciprocating engine generator sets as DG to meet peak load conditions on their system. The specifications of some of their DG projects are as follows:

- Hallock Substation, 18704 N. Krause Rd., Chillicothe, IL
- Eight reciprocating diesel engine generator sets
- Nominal capacity 1.6 MWe each, 12.8 MWe total
- Owned by AmerenCILCO
- Kickapoo Substation, 1321 Hickox Dr., West Lincoln, IL
- Eight reciprocating diesel engine generator sets
- Nominal capacity 1.6 MWe each, 12.8 MWe total
- Owned by Altorfer Inc.; power purchase agreement with AmerenCILCO, operating agreement provides for operations and maintenance (O&M)
- Tazewell Substation, 18704 N. Krause Rd., Chillicothe, IL
- Fourteen reciprocating diesel engine generator sets
- Nominal capacity 1.825 MWe each, 25.55 MWe total
- Owned by Altorfer Inc.; power purchase agreement with AmerenCILCO; operating agreement provides for O&M.

Although these DG power module facilities are primarily used as peaking facilities, they are also used to maintain system integrity in the event of an unanticipated outage at another AmerenCILCO generating station. They are unmanned and remotely operated from the company's Energy Control Center. They have proven to be a reliable and low cost option for the company to meet its peaking requirements. The power module sites were constructed at a cost of approximately $400/kW and have an operating cost of $75/MWh using diesel fuel at $0.85/gallon.

DG/CHP Projects

AmerenCILCO also has some successful CHP projects. A summary of two such DG/CHP projects are given below.

Indian Trails Cogeneration Plant

The Indian Trails Cogeneration Plant is owned and operated by AmerenCILCO. It is located on the property of MGP Ingredients of Illinois (MGP) in Pekin, Illinois and provides process steam to MGP and electricity to AmerenCILCO. The plant was constructed at a cost of $19,000,000 and went into full commercial operation in June 1995.

The plant consists of three ABB/Combustion Engineering natural gas-fired package steam boilers and one ABB STAL backpressure turbine-generator. Two of the boilers, boilers 1 and 2, are high-pressure superheat boilers rated at 185,000 lb/h of steam at 1250 psig and 900°F. Boiler 3 is a low-pressure boiler rated at 175,000 lb/h of steam at 175 psig and saturated temperature. Boilers 1 and 2 are normally in operation, with Boiler 3 on standby to insure maximum steam production reliability for MGP.

The high-pressure steam from boilers 1 and 2 passes through the ABB backpressure turbine-generator, which is rated at 21 MW. The steam leaving the turbine is at 175 psig and is desuperheated to 410°F to meet MGP's process steam requirements. The electricity produced goes to the AmerenCILCO grid to be used to meet utility system requirements.

The plant configuration provides significant operating efficiencies that benefit both MGP and AmerenCILCO. The Indian trails has an overall plant efficiency in excess of 80% and an electric heat rate of less than 5200 Btu/kWh. The construction of the Indian Trails by AmerenCILCO created an energy partnership with a valued customer. It allowed MGP to concentrate its financial and personnel resources on its core business. In turn, AmerenCILCO used its core business of producing energy to become an integral part of MGP's business, making AmerenCILCO more than just another vendor selling a product.

Medina Valley Cogen Plant

The Medina Valley Cogeneration Plant is owned and operated by AmerenCILCO. It is located on the property of Caterpillar and provides process steam and chilled water to Caterpillar, and electricity to AmerenCILCO. The plant was constructed at a cost of $64,000,000 and went into full commercial operation in September 2001.

The 40 MW electric generating plant consists of three natural gas-fired Solar Titan 130 model 18001S combustion turbines equipped with SoloNO_x (low NO_x) combustion systems manufactured by Caterpillar driving electric generators rated at 12.2 MW (gross generating capacity) each. There are also two Dresser-Rand steam turbine-generators with a total rated capacity of 8.9 MW.

The 410,000 #/h steam plant consists of three Energy Recovery International (ERI) VC-5-4816SH heat recovery steam generators (HRSGs) equipped with Coen low NO_x natural gas-fired duct burners and catalytic converters to reduce carbon monoxide (CO), rated at 109,000 lb/h at 600 psig each. There is also one Nebraska natural gas-fired steam generation boiler equipped with low NO_x burners, rated at 100,000 lb/h at 250 psig.

The plant configuration provides significant operating efficiencies that benefit both Caterpillar and AmerenCILCO. Medina Valley has an overall plant efficiency in excess of 70% and an electric heat rate of less than 6400 Btu/kWh. The construction of Medina Valley by AmerenCILCO created an energy partnership with a valued customer, whereby competitive electricity and steam prices were provided as well as greater operational flexibility, improved quality control in manufacturing, and improved steam reliability. It also allowed Caterpillar to concentrate its financial and personnel resources on its core business. In turn, AmerenCILCO used its core business of producing energy to become an integral part of Caterpillars business, strengthening its ties with a major customer as well as adding additional efficient-generating capacity, and improving air quality (399 fewer tons pollutants/year).

CONCLUSIONS

Distributed generation and DG/CHP should, must, and will happen. The benefits to all parties when CHP is utilized are too much to ignore. Therefore, the question becomes who should do it, not should it be done.

If management is behind the project, the engineering and maintenance staff is capable, and financing is available, the project should be done in-house. Maximum economic benefits would result. However, the user must commit to this project.

If any of the above is not true, the best approach for the facility or plant is to seek the help of an ESCO. It is important that the ESCO chosen must be fully equipped with knowledge and experience in all of the technologies and be able to provide the financing package. Such ESCOs do exist today, but some of them need a better understanding of the different technologies required, as well as insuring that the DG/CHP project is successfully completed and turned over to a facility that can operate it and maintain it. Commissioning and baton passing must be part of the contract.

The authors believe there is a tremendous opportunity for local utility ESCOs to successfully participate in this movement to DG, and particularly to the use of CHP. A utility-affiliated ESCO, properly equipped as we have defined it, can do these projects, and can do them successfully. The local utility ESCO has entries with local facilities and plants that few other ESCOs have. If the utility can exploit this opportunity, they have the chance to help many facilities and to help themselves in the process. This is a true win–win opportunity for the utility affiliated ESCO. All utilities should be ready to fill this need or recognize that they will likely lose market share.

REFERENCES

1. Capehart, B.L., Turner, W.C., Kennedy, W.J., Eds.; *The Guide to Energy Management*, 5th Ed.; Fairmont Press: Lilburn, GA, 2006 (Chapter 14).
2. Turner, W.C. Ed.; *Energy Management Handbook*, 6th Ed.; Fairmont Press: Lilburn, GA, 2006 (Chapter 7).
3. Petchers, N. Ed.; *Combined Heating, Cooling and Power Handbook*, Fairmont Press: Lilburn, GA, 2003.

District Cooling Systems

Susanna S. Hanson
Global Applied Systems, Trane Commercial Systems, La Crosse, Wisconsin, U.S.A.

Abstract
Seemingly small efficiency improvements to traditionally designed chiller plants multiply to create impressive savings in district cooling plants. The size and scope of these larger projects make the benefits more obvious and very valuable. For example, the United Arab Emirates is fully embracing district cooling for accommodating its tremendous growth, rather than using individual cooling systems in each building. It has been estimated that 75% of the energy used in Dubai, U.A.E. is for cooling. By using district cooling, Dubai planners expect to reduce the amount of electrical energy used by 40%. More than 30% of the Dubai market is selecting pre-engineered, packaged chilled water systems, many of which employ series chillers, variable-primary distribution design, and other concepts discussed in this article.

INTRODUCTION

District cooling was inspired by district heating plants, which became popular in the 19th century for providing steam to a number of buildings and often to entire portions of cities. Recent district cooling projects exceed 100,000 tons, typically in urban areas with dense, simultaneous development. One project of this size can, all by itself, magnify an incremental efficiency improvement into nearly 10 MW and offset $10 million or more in power generation capacity.

These ideas can naturally be used on smaller projects, where the benefits are less immediately obvious. To illustrate the energy and cost savings potential, this entry includes an analysis from a 10,500-ton convention center. The key design elements used on these projects violate traditional rules of thumb and allow the system designer to focus on the energy consumed by the entire chilled water system.

MOVING AWAY FROM CHILLER-ENERGY-ONLY ANALYSES

In the last 40 years, water-cooled chiller efficiency expressed in coefficient of performance (COP) has improved from 4.0 to over 7.0. Cooling tower and pump energy has remained largely flat over the same period. The result is that towers and pumps account for a much higher percentage of system energy use.

Because of these dramatic improvements, it is tempting to go after chiller savings in new plants and chiller replacements. After all, the chiller usually has the biggest motor in a facility. Many chiller plant designs continue to use flow rates and temperatures selected to maximize chiller efficiency and ignore the system effects of those decisions. Chilled water system designs frequently default to using the flow rates and temperatures used for the rating tests developed by the Air Conditioning and Refrigeration Institute (ARI) standards 550/590 for vapor compression chillers[1] and ARI 560 for absorption chillers.

While these benchmarks provide requirements for testing and rating chillers under multiple rating conditions, they are not intended to prescribe the proper or optimal flow rates or temperature differentials for any particular system. As component efficiency and customer requirements change, these standard rating conditions are seldom the optimal conditions for a real system. There is great latitude in selecting flow rates, temperatures, and temperature differences.

Today, an equal price centrifugal chiller can be selected for less condenser flow, with no loss in efficiency. At the same time, chilled water temperatures are dropping. Pump savings usually exceed chiller efficiency losses. Larger chilled water distribution systems will realize higher savings. Analyze the entire system to define the right design conditions.

MARKET TRENDING TOWARD HIGHER LIFT

Chiller efficiency is dependent on several variables—capacity (tons) and lift (chiller internal differential temperature) are two of them. Lift is the difference between the refrigerant pressures in the evaporator and condenser, and it can be approximated using the difference between the water temperatures leaving the evaporator and condenser.

Designers are converging on higher-lift systems for a variety of reasons; e.g., more extreme climates, system

Keywords: Chillers; Series chillers; Counterflow; Variable-primary pumping; District cooling; Efficient chiller plant design; Chiller lift; Part load efficiency.

optimization, replacement considerations, thermal storage, and heat recovery. Higher lift creates solutions for engineering problems and leverages chiller capabilities.

Chiller plant design for U.S. conditions usually calls for a maximum tower-leaving temperature of 85°F. Conventional designs used 44°F/54°F evaporators and 85°F/95°F condensers—easy on the chiller (and the system designer). But higher ambient wet bulb conditions are common in the Middle East and China. Many parts of China call for 89.6°F design condenser/tower water, and parts of the Middle East design for 94°F. As the cooling markets grow in areas with extreme weather, technology providers must find better ways to deliver cooling at the higher lift conditions.

A reduced flow rate in the condenser saves tower energy and condenser pump energy. In the replacement and renovation market, the resulting increase in delta-T (tower range) delivers the same capacity with a smaller cooling tower, or more capacity with the existing cooling tower.

Chilled water temperatures below 40°F allow system designers to implement a larger chilled water delta-T. The result is less costly chilled water distribution and more effective cooling and dehumidification. Besides reducing water flow rates, colder chilled water gives higher delta-Ts at the chiller and better utilization of the chillers.

Thermal storage is also becoming more prevalent. For stratified chilled water storage tanks, 39°F water is the magic number, because it corresponds to the maximum density of water and keeps the charged section below the thermocline.[2]

Higher condensing temperatures provide more useful heat to recover from the condenser water. Some commercial building codes require condenser heat recovery for applications with simultaneous heating and cooling.

A proper chiller selection can deliver these conditions with little or no extra cost. It is possible to increase lift while increasing overall chiller plant efficiency by using a series-counterflow chiller arrangement.

SERIES CHILLERS

In their heyday of the early 1960s, series chillers were widely used in government buildings in Washington, D.C. Series arrangements were necessary for perimeter induction cooling systems, which supply cold primary air to the space and require colder water from the chiller. As previously discussed, chillers in the 1960s had a COP of about 4.0, with high-flow (velocity) smooth-bore tubes, low tube counts, and one-pass evaporators to reduce pressure drop. The most efficient centrifugal chillers today have a COP of more than 7.0—more than 75% higher than chillers used in these early series chiller plants.

Designers virtually stopped using series arrangements in the 1970s because variable air volume (VAV) systems made the colder chilled water used for induction systems unnecessary. Given the chiller's relatively low efficiency by today's standards, it made sense to raise temperatures and save chiller energy. Variable air volume systems were widely adopted because they saved energy and adapted to unknown cooling loads. Variable air volume systems are still the most popular choice for comfort cooling applications.

Induction systems are virtually nonexistent in 2006, but the dramatically improved centrifugal chiller efficiencies are driving resurgence in series chiller arrangements. Series chiller plants offer energy and first cost savings, even in VAV systems.

Using multiple-stage centrifugal chillers and putting chillers in series creates higher lift more easily and more efficiently, with a more rigorous, stable operating profile. In contrast, chillers lined up in parallel must each create the coldest water required for the entire system while rejecting heat to the warmest condensing temperature.

While series arrangements of chiller evaporators have been used in many applications,[3,4] series condenser arrangements are less common. A series-counterflow chilled water plant design arranges the evaporators in series, but also arranges the condensers in series using a counterflow configuration. Counterflow means that the condenser water and the evaporator water flow in opposite directions.

Chiller plants with series evaporators and parallel condensers aren't recognizing the highest efficiency gains because the chiller producing the coldest chilled water must create more lift and therefore do more work. By arranging the condensers in series counterflow, the lift of each compressor is nearly the same. The result is a pair of chillers working together to create high lift while increasing overall plant efficiency.

VARIABLE FLOW

A relatively new concept, the variable-primary system, has removed one barrier to series-chiller plant design. Series chillers often have a higher pressure drop. In traditional primary–secondary systems, constant flow through the chillers equals constant pressure drop and constant pump energy, so the series chillers' higher pressure drop results in higher pump energy all the time. Varying the flow through series chillers eliminates the penalty for much of the operating hours.

Variable-primary systems send variable amounts of water flow through the chillers to reduce the pumping energy and enable the delta-T seen by the chiller to remain equal to the system delta-T. Because pump energy is approximately proportional to the cube of the flow (subject to losses through the distribution system), even small flow reductions are valuable.

Consider the following ideal relationship between flow and energy in a pumping system:

$$\text{Pump Energy} \propto \text{Flow}^3$$

For a system using 80% of the design flow—a modest 20% reduction—the energy required is reduced by nearly 50%. In turn, a more aggressive 50% reduction in flow is an 87% reduction in power.

Varying the flow through series chillers reduces the total operating cost, despite an increased pressure drop at design conditions. Improved tube designs and extensive testing in manufacturers' testing labs have cut minimum water velocities in half, leading to better turndown for all chillers. The additional implicit turndown capability of the series configuration enables further pumping energy savings. Single-pass evaporators and condensers, when practical, reduce water pressure drop and pumping costs compared to two- or three-pass configurations.

VARIABLE FLOW COMBATS THE INCREDIBLE SHRINKING CHILLER

So far, we have only discussed efficiency and energy savings. But what about the installed costs associated with chiller-water system design? Any chiller, big or small, is a significant capital investment. This is one reason most chillers are factory tested for capacity. Cooling capacity is the product of chilled water flow and chilled water delta-T. For the same capacity, as gallons per minute goes up, delta-T goes down.

Chilled water system design has a unique and sometimes dramatic effect on chiller capacity. Primary–secondary chilled water systems shrink chiller capacities, while variable-primary chilled water systems help chiller capacities expand. At times, this chiller capacity expansion can exceed nominal chiller capacity.

Due to this variation in chiller capacity, primary–secondary chilled water systems have excessive chiller starts as well as higher chiller run hours, because more chillers will be operating than necessary. Variable-primary chilled water systems have fewer chiller starts and lower chiller run hours, by squeezing more capacity out of the operating chillers. The plant controller can delay the operation of an additional chiller by increasing the flow through the operating chillers, thus increasing the chiller capacity. Operating the fewest number of chillers possible is a well-known energy optimization strategy.

It is the design of the chilled water system that causes this change in net chiller capacity, not the chiller itself. Chillers are constant flow devices when employed in a primary–secondary chilled water system. A bypass pipe, commonly called a decoupler, serves as a bridge between the primary loop serving only the chillers and the secondary loop distributing chilled water throughout the building to all the air handling units. Because the chiller is a constant flow device, we must subject the chiller to a smaller chilled water delta-T in order to unload the chiller. The decoupler pipe mixes surplus-chilled water with water returning from the load to produce cooler return water for the chiller. But what if some event other than the decoupler lowers the temperature of the returning chilled water? This reduction of chiller-entering water temperature unloads the chiller—even when we don't want the chiller unloaded.

In a primary–secondary system, additional chillers are sequenced on when the demand for chilled water exceeds the constant flow capacity of each chiller. To illustrate, consider three equally sized chillers in a primary–secondary system. Each chiller is sized for 500 tons and 750 gpm of chilled water (1.5 gpm per ton, or a 16°F chilled water delta-T). When the chilled water distribution system demands more than 750 gpm, two chillers must operate. When the chilled water system demand exceeds 1500 gpm, all three chillers must be on. These chillers will produce their full 500-ton cooling capacity only when the cooling coils create a full 16°F water-temperature rise. If the cooling coils collectively can only produce a 12°F water-temperature rise, the 500-ton chillers can only produce 375 tons. 125 tons of installed cooling capacity is lost. This inability of the chilled water distribution system to achieve the design chilled-water-temperature rise is called "low delta-T syndrome."

There are several contributors to low delta-T syndrome, but the three greatest offenders are excess distribution pump head, three-way control valves, and chilled water reset. Three-way control valves allow cold chilled water to bypass the cooling coil. The bypassed water is dumped into the return chilled water line, diluting the return chilled water. Any high water-temperature rise created by the cooling coil is destroyed.

The chilled water distribution pump is variable speed, producing more pressure when more chilled water flow is required, and producing less pressure when less chilled water flow is required. The speed of the pump is controlled by sensing the available pressure at the end of the chilled water distribution loop. If this pump creates excess pressure, even two-way control valves will have a difficult time reducing chilled water flow through the cooling coil. When the cooling coil receives excess flow, there is not enough heat in the air to adequately warm the water. Again, water-temperature rise is hampered.

Perhaps the most insidious destroyer of water-temperature rise is chilled water reset. The chiller will consume less energy when it produces warmer chilled water, but chiller energy savings may be dwarfed by the additional pump energy consumption required for delivering warmer chilled water. Coils use less chilled water when it is delivered at a colder temperature. As the chilled water

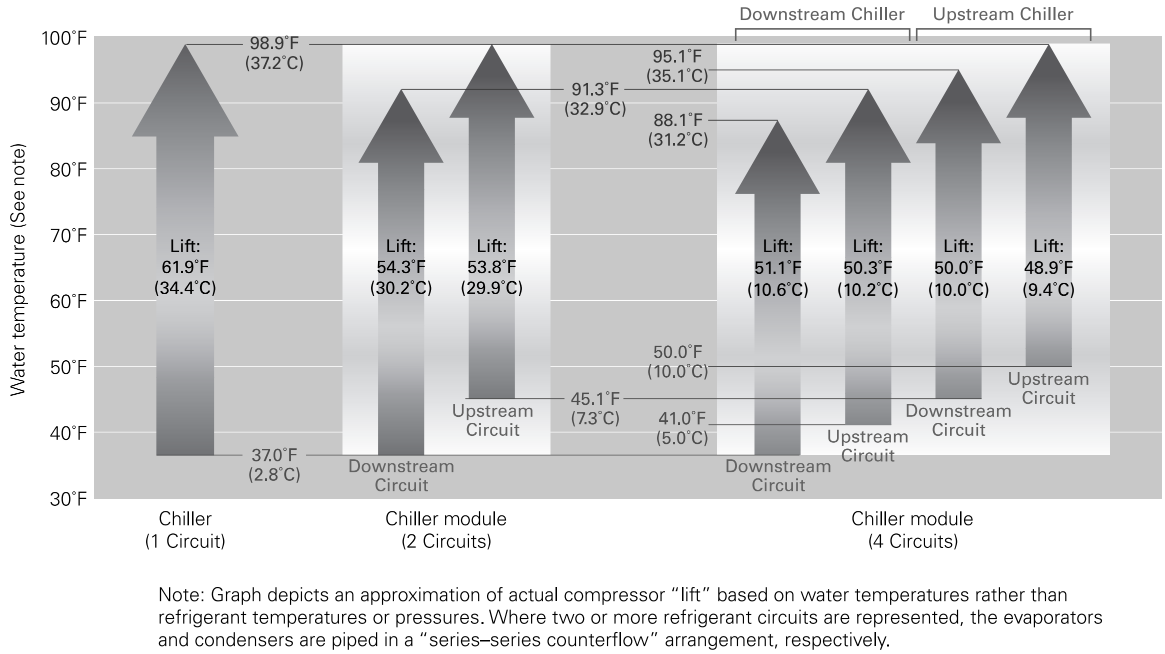


Fig. 1 Series counterflow chiller arrangement equalizes lift performed by each compressor, minimizing the energy needed to create high lift.

temperature is set upwards, each coil will demand more water to meet the same cooling load.

CHILLER PLANT DESIGN ENERGY COMPARISON

The following example is for a design created initially for the Washington DC Convention Center. Each pair of series-counterflow chillers (assuming multiple-stage compressors on each circuit) has eight to twelve stages of compression equally sharing the load (Fig. 1). The chiller module depicted in Fig. 2 created series-pair efficiencies of 0.445 kW per ton (7.8 COP) at standard ARI rating conditions.

Fig. 3 shows the component and system energy use of various parallel and series chiller configurations using variable evaporator flow with reduced condenser water flow. The series–series counterflow arrangement for the chillers reduces the chiller energy to compensate for additional pump energy. In the case of this particular installation, series–series counterflow saved $1.4 million in lifecycle costs over the parallel–parallel alternative. The low-cost alternative used six electric centrifugal chillers with dual refrigeration circuits, 2 gpm of condenser water per ton of cooling, piped in a "series evaporator-series condenser" arrangement. It also used 1040 kW less than the parallel–parallel configuration.[5]

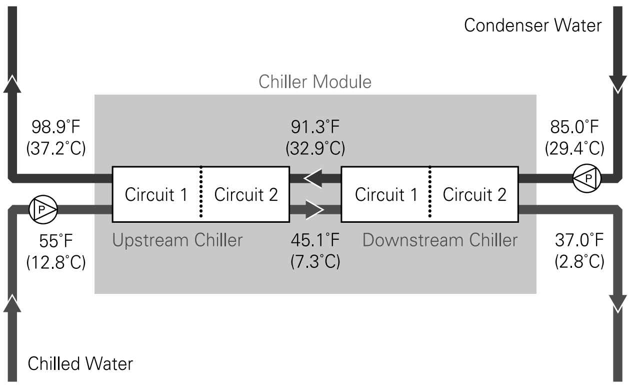


Fig. 2 Module with dual-circuit chillers in series provides 8–12 stages of compression and uses 0.445 kW per ton for the chillers at standard ARI rating conditions.

Arrangement		Chillers*		Evaporator				Condenser				Cooling Towers		System	
Evaporator	Condenser	Units/ modules	Compressor efficiency kW/ton	Flow gpm	ΔP Feet of Water	Number of pumps	Power per pump kW	Flow gpm	ΔP Feet of Water	Number of Pumps	Power per pump kW	Number of cells	Power per cell kW	Total power kW	Life-cycle cost $USD
Parallel	Parallel	5/5	0.649	2,800	3.26	5	2.18	4,200	3.66	5	3.67	8	60	7324	18,836,302
Parallel	Parallel	6/6	0.618	2,333	4.18	6	2.33	3,500	3.53	6	2.95	8	60	7001	18,076,391
Series	Series-Counterflow (1.5 gpm/ton)	6/3	0.560	4,667	17.96	3	19.99	5,250	14.8	3	18.54	8	48	6379	16,819,167
Series	Series-Counterflow (2.0 gpm/ton)	6/3	0.535	4,667	17.96	3	19.99	7,000	25.2	3	42.08	8	60	6284	16,656,947
Series	Parallel (2.0 gpm/ton)	6/3	0.555	4,667	17.96	3	19.99	3,500	3.53	6	2.95	8	60	6385	16,888,493

* The chillers represented in this table all have dual refrigerant circuits. The full analysis included single refrigerant circuit chillers at various flow rates and efficiencies.

Fig. 3 Projected energy-use and lifecycle costs for series and parallel chiller configurations.

DESIGN PARAMETERS

A larger-than-conventional difference between the entering and leaving chilled water temperatures permits a lower flow rate, reducing the initial costs for distributing the chilled water (pumps, piping) in central chilled water plants. Smaller pipes and pumps can then be used to satisfy the same capacity.

Because supplying colder chilled water requires more power from the chillers, the cost savings from reducing the pumping power and pipe size and installation must offset the chiller power increase. Chiller designs and controls have improved to the point at which producing 37°F water no longer causes concern for freezing evaporator tubes. Experience shows that fast, accurate chiller controls and algorithms can safely accommodate temperatures as low as 34°F without the addition of antifreeze.

- Entering-chiller water temperature: 55°F.
- Leaving-chiller water temperature: 37°F.
- Evaporator flow rate/capacity: 1.33 gpm/ton.

Many plant configurations are possible:

- Both evaporators and condensers in parallel.
- Evaporators in series and condensers in parallel.
- Both evaporators and condensers in series.

At design conditions:

- Chilled water enters the upstream chiller at 55°F and exits at 45.1°F.
- Chilled water enters the downstream chiller at 45.1°F and exits at 37°F.
- Condenser water enters the downstream chiller at 85°F and exits at 91.3°F.
- Condenser water enters the upstream chiller at 91.3°F and exits at 98.9°F.

FULL-LOAD EFFICIENCY IMPROVEMENT

The series–series counterflow arrangement yields the lowest full-load chiller power (about 14% lower than the parallel–parallel configuration) (Fig. 3). The dramatic reduction in chiller power occurs because the upstream chiller in the series–series counterflow arrangement operates at a higher chilled water temperature, which means that the refrigerant temperature and refrigerant pressure in the evaporator are also higher in the upstream machine. The downstream chiller "sees" a lower condenser-leaving water temperature—and therefore has a lower condenser refrigerant pressure—than it would in a plant with the chiller condensers arranged in parallel. Fig. 1 illustrates the concept of reduced lift using the design parameters for this chilled water plant. Chiller power can be reduced by decreasing compressor "lift." In this example, the difference in average lift at design is nearly 13%.

$$1 - \frac{(54.3 + 53.8)/2}{61.9} = 0.126$$

Now, consider a series–series counterflow arrangement of two dual-circuited chillers. Because each of the chillers in this design has two refrigeration circuits, the reduced lift effect is multiplied. Instead of two lifts, there are four. The difference in average lift at design for the system with four independent refrigeration circuits in a series–series counterflow arrangement exceeds 19%.

$$1 - \frac{(51.1 + 50.3 + 50.0 + 48.9)/4}{61.9} = 0.191$$

At the design conditions defined for the system, chiller performance is well above the 6.1 COP requirement set by ANSI/ASHRAE/IESNA Standard 90.1-2004, Energy Standard for Buildings Except Low-Rise Residential Buildings. At standard ARI rating conditions, each chiller module would operate with an efficiency of 0.445 kW/ton.

PART-LOAD EFFICIENCY IMPROVEMENT

The reduction in lift provided by the series–series counterflow arrangement also occurs at part-load conditions. The temperature of the water leaving the evaporator of the upstream chiller is always warmer than the system water, and the temperature of the water leaving the condenser of the downstream chiller is always cooler than the system water. The upstream chiller does not need to perform the same amount of cooling as the downstream chiller. The benefit comes from the upstream chiller's ability to produce chilled water at an elevated temperature.

SAVINGS AMPLIFIED BY POWER INFRASTRUCTURE

While the chiller performance is remarkable with this design, the performance conditions for this application were carefully selected to optimize the overall energy consumption of the entire chilled water plant. Series chiller configurations are not just about the chiller or the pump savings, but reduced electrical infrastructure requirements and environmental impact. The previous example was a 10,500-ton plant. Consider the reduction in power generation requirements when multiplied tenfold. In a 100,000-ton cooling and power infrastructure project, increasingly common in the Middle East and China, our example's 1040-kW reduction blossoms into nearly 10 MW. A conservative estimate for the cost of the generation equipment is $1000 per kW. Using that round budget number, series chiller configuration and low-flow, low-temperature conditions could save $10 million or more in power generation equipment on a 100,000-ton project.

PACKAGED SOLUTIONS MINIMIZE COMPLEXITY AND RISK

Frequent users of these concepts are packaged chiller plant manufacturers that are pre-engineering and packaging series-counterflow chillers with built-in optimization controls. Packaged chiller plants utilizing series-counterflow chiller arrangements are currently available in sizes up to 8000 tons. The chillers in these packaged chiller plants can be factory-performance tested in accordance with ARI procedures prior to shipment from the chiller manufacturers' facility.

Packaging companies and astute engineers have put two series-counterflow chillers in series with each other—essentially creating a 4-chiller series module (Fig. 2). These solutions enhance the thermodynamic benefit created by series-counterflow chillers while minimizing complexity and risk for the system owner and operator.

CONTROL STRATEGIES

How the plant should respond to varying system conditions is a topic for discussion with the design engineer, plant owner, and plant operators. For example, if the entering-chiller water temperature is not reaching design conditions, the operators could:

1. Increase pump speed or turn on more pumps to increase flow rates and more fully load the active chillers.
2. Reset the setpoints of the upstream chillers to 55% of the total temperature difference. Lowering the setpoint of the upstream chillers as the result of a drop in entering-chiller water temperature lessens the benefit of reduced lift. However, the upstream chillers will always run at a higher evaporator pressure than the downstream chillers, which saves energy consumption and costs.
3. Address chiller sequencing in the context of the system options, variable or constant chiller flow, extra pumps, or other concerns. It might be most cost effective to use a startup strategy that fully loads one chiller module and then activates the remaining chillers in modules (pairs). Activating the upstream chiller and operating it at the higher water temperature takes advantage of all of the available heat transfer surface area without increasing the energy consumed by ancillary equipment.

Dem–Dry

USEFUL REDUNDANCY

Large chiller plants can be more adaptive and efficient with multiple chillers rather than fewer large, field-erected chillers. In plants with more chillers, redundancy is easily created through parallel banks of upstream and downstream chillers. Different combinations of upstream and downstream chillers can meet the load, so if one chiller is being serviced, its duty can be spread out to the other chillers. The same is true for pumps, which do not have to be sequenced with the chillers.

EFFICIENCY, FLEXIBILITY IN SMALLER PLANTS AND RETROFITS

The benefits of low flow, low temperature, and high efficiency apply to other types of chillers as well. Smaller, noncentrifugal chillers can benefit proportionately more under these conditions when placed in series.

Helical-rotary chillers are sensitive to increased lift and decreased condenser water flow. Absorption chillers struggle to make water colder than 40°F, and their cooling capacity increases when placed upstream. Both can be put upstream in the sidestream position for reduced first cost and higher efficiency. Reusing existing, older, less efficient chillers upstream is also an interesting option to explore. These sidestream configurations combine the benefits of series and parallel chillers while isolating some chillers from water flow variations.

CONCLUSION

As chiller efficiencies continue to improve, district energy and central plant designers can optimize the entire system to achieve even lower costs of ownership. Owners can expect more first cost and energy savings from low-flow, low-temperature and highly efficient chiller configurations. The unique benefits and flexibility of series chiller plant designs with variable-primary pumping arrangements include lower overall chilled water system operating costs, reduced emissions, and improved environmental responsibility.

REFERENCES

1. *Standard for water chilling packages using the vapor compressor cycle*, ARI Standard 550/590-2003, Air-Conditioning and Refrigeration Institute, Arlington, VA, 2003.
2. *ASHRAE Handbook-HVAC Systems and Equipment*, Chapter 11, District Heating and Cooling, 2000; 11.1-11.34.
3. Coad, W.J. A fundamental perspective on chilled water systems. Heat Piping Air Cond. **1998**, *August*, 59–66.
4. Trane, *Multiple-Chiller-System Design and Control*, SYS-APM001-EN, American Standard Inc., 2001; 58–59.
5. Groenke, S.; Schwedler, M. ASHRAE J. **2002**, *44* (6), 23–31.

BIBLIOGRAPHY

1. *ASHRAE Handbook-HVAC Systems and Equipment Handbook*, Chapter 12, Hydronic Heating and Cooling System Design, 12.1-12.18.
2. Avery, G. Improving the efficiency of chilled-water plants. ASHRAE J. **2001**, *43* (5), 14–18.
3. Avery, G. Controlling chillers in variable flow systems. ASHRAE J. **1998**, *40* (2), 42–45.
4. Bahnfleth, W.P.; Peyer, E. Comparative analysis of variable and constant primary-flow chilled-water-plant performance. HPAC Eng. **2001**, *April*, 50.
5. Demirchian, G.H.; Maragareci, M.A. *The Benefits of Higher Condenser Water DT at Logan International Airport central chilled water plant*, IDEA 88th Annual conference Proceeding, 1997; 291–300.
6. Hartman, T. All-variable speed centrifugal chiller plants. ASHRAE J. **2001**, *43* (9), 43–51.
7. Kelly, D.W.; Chan, T. Optimizing chilled water plants. Heat Piping Air Cond. **1999**, *January*, 145–147.
8. Kirsner, W. The demise of the primary–secondary pumping paradigm for chilled water plant design. Heat Piping Air Cond. **1996**, *November*, 73–78.
9. Luther, K. Applying variable volume pumping. Heat Piping Air Cond. **1998**, *October*, 53–58.
10. Rishel, J.B. *HVAC Pump Handbook*; McGraw Hill: New York, 1996; 109–112.
11. Schwedler, M.; Bradley, B. Variable-primary-flow systems: An idea for chilled-water plants the time of which has come. Heat Piping Air Cond. **2000**, *April*, 41–44.
12. Taylor, S. Primary-only vs primary–secondary variable flow chilled water systems. ASHRAE J. **2002**, *44* (2), 25–29.
13. Waltz, J.P. Don't ignore variable flow. Contracting Bus. **1997**, *July*, 133–144.

District Energy Systems

Ibrahim Dincer
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology (UOIT), Oshawa, Ontario, Canada

Arif Hepbasli
Department of Mechanical Engineering, Faculty of Engineering, Ege University, Izmir, Turkey

Abstract

This entry presents some historical background on district heating and cooling along with cogeneration and geothermal applications, and discusses some technical, economical, environmental, and sustainability aspects of geothermal energy and performance evaluation tools in terms of energy and exergy analyses for district heating systems. Case studies are also presented to highlight the importance of exergy use as a potential tool for system analysis, design, and improvement.

Dem–Dry

NOMENCLATURE

COP	coefficient of performance
E	energy
Ex	exergy
f	figure of merit
h	specific enthalpy
Q	heat interaction
R	energy grade function
s	specific entropy
T	temperature
W	shaft work
η	energy efficiency
τ	exergetic temperature factor
ψ	exergy efficiency

Subscripts

C	cooling
ch	chiller
d	natural direct discharge
dest	destruction
DH	district heating
elec	electrical
equiv	equivalent
f	fuel
gen	generation
H	heating
HE	heat exchanger
heat	heat
in	inlet
net	net
o	environmental state
out	outlet
r	reinjected
sys	system
tot	total
UC	user cooling
UH	user heating

Superscripts

·	rate with respect to time
CHP	combined heat and power (cogeneration)
r	room
s	space
u	user
w	water

Abbreviations

CHP	combined heat and power
DC	district cooling
DES	district energy system
DH	district heating
DHC	district heating and cooling
GDHS	geothermal district heating system
IBGF	Izmir–Balcova geothermal field
IEA	International Energy Agency
IPCC	Intergovernmental Panel on Climate Change

Keywords: District heating; District cooling; Cogeneration; Efficiency; Energy; Exergy; Environment; Sustainability; Performance; System.

INTRODUCTION

District energy systems (DESs) can utilize a wide range of energy resources, ranging from fossil fuels to renewable energy to waste heat. They are sometimes called community energy systems because, by linking a community's energy users, DESs maximize efficiency and provide opportunities to connect generators of waste energy (e.g., electric power plants or industrial facilities) with consumers who can use that energy. The heat recovered through district energy can be used for heating or can be converted to cooling using absorption chillers or steam turbine drive chillers.

District energy system cover both district heating (DH) and district cooling (DC), and distribute steam, hot water, and chilled water from a central plant to individual buildings through a network of pipes. District energy systems provide space heating, air conditioning, domestic hot water, and/or industrial process energy, and often also cogenerate electricity. With district energy, boilers and chillers in individual buildings are no longer being required. District energy is considered an attractive, more efficient, and more environmentally friendly way to reduce energy consumption.

A basic DH system consists of three main parts, as shown in Fig. 1a, and for DC the flow chart becomes somewhat different, as shown in Fig. 1b. An example is that the DC system uses hot water produced from the DH system to operate an absorption refrigerator installed in the substation of the consumers' building instead of the conventional vapor-compression refrigeration system operated by electricity.

Note that storage of chilled water or ice is an integral part of many DC systems. Storage allows cooling energy to be generated at night for use during the hottest part of the day, thereby helping manage the demand for electricity and reducing the need to build power plants.[1]

District heating and cooling (DHC) systems can provide other environmental and economic benefits, including:

- Reduced local and regional air pollution
- Increased opportunities to use ozone-friendly cooling technologies
- Infrastructure upgrades and development that will provide new jobs
- Enhanced opportunities for electric peak reduction through chilled water or ice storage
- Increased fuel flexibility
- Better energy security

In fact, the Intergovernmental Panel on Climate Change (IPCC) has identified cogeneration/DHC as a key greenhouse-gas reduction measure, and the European Commission has been developing a European Union cogeneration/DHC strategy.

District heating and cooling potential can be realized through policies and measures to increase awareness and knowledge; recognize the environmental benefits of district energy in air-quality regulation; encourage investment; and facilitate increased use of district energy in government, public, commercial, industrial, and residential buildings.

During the past few decades there have been various key initiatives taken by major energy organizations (e.g., International Energy Agency (IEA), U.S. Department of Energy, Natural Resources Canada, etc.) on the implementation of DHC all over the world as one of the most significant ways to:

- Maximize the efficiency of the electricity generation process by providing a means to use the waste heat, saving energy while displacing the need for further heat-generating plants
- Share heat loads, thereby using plants more effectively and efficiently
- Achieve fuel flexibility and provide opportunities for introduction of renewable sources of energy as well as cogeneration and industrial waste heat

Furthermore, the IEA recently developed a strategic document[2] as an implementing agreement on DHC including the integration of combined heat and power (CHP), focusing on:

- Integration of energy-efficient and renewable energy systems for limited emissions of greenhouse gases

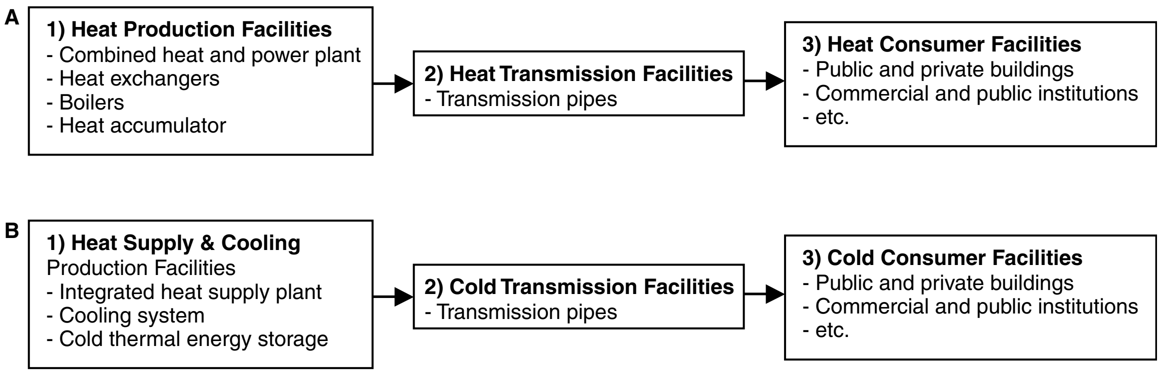


Fig. 1 A basic flow chart for (A) DH application and (B) DC application.

Dem–Dry

- Community system integration and optimization, and use of waste thermal energy, renewable energy and CHP, for a better environment and sustainability
- Reliability, robustness, and energy security for effective maintenance and management of buildings
- Advanced technologies for improved system integration, including information systems and controls
- Dissemination and deployment for rapid change toward energy efficiency and sustainability

In this entry we present some historical background on DHC systems and applications, and discuss some technical, economical, environmental, and sustainability aspects of these systems and their performance evaluations tools in terms of energy and exergy efficiencies. A case study is also presented to highlight the importance of exergy use as a potential tool for system analysis, design, and improvement.

COGENERATION AS A KEY PART OF DISTRICT HEATING AND COOLING

Cogeneration, also referred to as CHP, is the simultaneous sequential production of electrical and thermal energy from a single fuel. During the past couple of decades, cogeneration has become an attractive and practical proposition for a wide range of thermal applications, including DHC. Some examples are the process industries (pharmaceuticals, paper and board, cement, food, textile, etc.); commercial, government, and public-sector buildings (hotels, hospitals, swimming pools, universities, airports, offices, etc.); and DHC schemes. Fig. 2 shows a comparison of conventional power systems and cogeneration systems. The main drawback in the conventional system is the amount of intensive heat losses, resulting in a drastic drop in efficiency. The key question is how to overcome this and make the system more efficient. The answer is clear: by cogeneration. In this regard, we minimize the heat losses and increase the efficiency, and provide the opportunity to supply heat to various applications and facilities. The overall thermal efficiency of the system is the percentage of the fuel converted to electricity plus the percent of fuel converted to useful thermal energy. Typically, cogeneration systems have overall efficiencies ranging from 65 up to 90%, respectively.

The key point here is that the heat rejected from one process is used for another process, which makes the system more efficient compared with the independent production of both electricity and thermal energy. Here, the thermal energy can be used in DH and/or DC applications. Heating applications basically include generation of steam or hot water. Cooling applications basically require the use of absorption chillers that convert heat to cooling. Numerous advanced technologies are available to achieve cogeneration, but the system requires an electricity generator and a heat-recovery system for full functioning.

Cogeneration has been widely adopted in many European countries for use in industrial, commercial/institutional, and residential applications. It currently represents 10% of all European electricity production and more than 30% of electricity production in Finland, Denmark, and the Netherlands. Within Canada, however, cogeneration represents just over 6% of national electricity production.[3] This relatively lower penetration is attributed to Canada's historically low energy prices and electric-utility policies on the provision of backup power and the sale of surplus electricity. Despite these conditions, cogeneration has been adopted in some industrial applications, notably the pulp-and-paper and chemical products sectors, where a large demand for both heat and electricity exists. Several classical technologies currently are available for cogeneration, such as steam turbines, gas turbines, combined cycle (both steam and gas) turbines, and reciprocating engines (gas and diesel). In addition, there has been increasing interest in using some new technologies—namely, fuel cells, microturbines, and Stirling engines. Note that heat output from

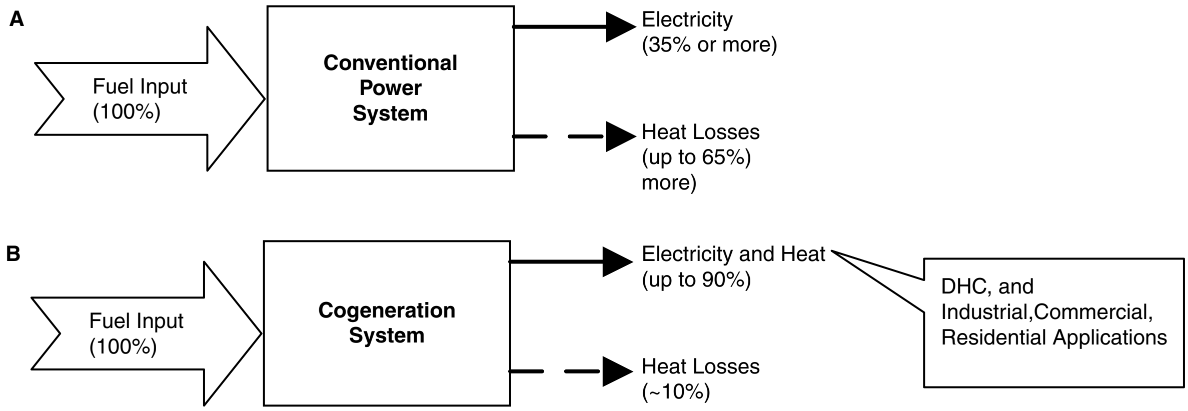


Fig. 2 Illustration of (A) conventional power system and (B) cogeneration system.

Table 1 Main characteristics and technical aspects of cogeneration systems

Technology	Fuel type	Capacity (MW_e)	Electrical efficiency (%)	Overall efficiency (%)	Average capital cost (US$/$kW_e$)	Average maintenance cost (US$/kWh)
Steam turbine	Any	0.5–500	7–20	60–80	900–1800	0.0027
Gas turbine	Gaseous and liquid fuels	0.25–50 or more	25–42	65–87	400–850	0.004–0.009
Combined cycle	Gaseous and liquid fuels	3–300 or more	35–55	73–90	400–850	0.004–0.009
Reciprocating engines	Gaseous and liquid fuels	0.003–20	25–45	65–92	300–1450	0.007–0.014
Micro turbines	Gaseous and liquid fuels	-	15–30	60–85	600–850	<0.006–0.01
Fuel cells	Gaseous and liquid fuels	0.003–3 or more	35–50	80–90	—	—
Stirling engines	Gaseous and liquid fuels	0.003–1.5	~40	65–85	—	—

Source: From United Nations Environment Programme (see Ref. 4).

the system varies greatly depending on the system type. The output can range from high-pressure, high-temperature (e.g., 500°C–600°C) steam to hot water (e.g., 90°C). High-pressure, high-temperature steam is considered to be high-quality thermal output because it can meet most industrial-process needs. Hot water is considered to be a low-quality thermal output because it can be used for only a limited number of DHC applications.

Cogeneration can be based on a wide variety of fuels, and individual installations may be designed to accept more than one fuel. Although solid, liquid, or gaseous fossil fuels currently dominate, cogeneration from biomass fuels is becoming increasingly important. Sometimes, fuels are used that otherwise would constitute waste (e.g., refinery gases, landfill gas, agricultural waste, and forest residues). These substances increase the cost efficiency of cogeneration.[4] Table 1 gives a comprehensive summary of cogeneration technologies in terms of fuel type, capacity, efficiency, average capital cost, maintenance cost, etc.

HISTORY OF DISTRICT HEATING AND COOLING

The oldest DH system was created in the early 14th century in the village of Chaudes-Aigues Cantal in France. This system distributed warm water through wooden pipes and is still in use today. The first commercial DH system was created by Birdsill Holly in Lockport, New York in 1877. In this system the boiler was used as the central heat source; it supplied a loop consisting of steam pipes, radiators, and even condensate-return lines. Later, the system attracted a dozen customers. Only three years later, it served several factories as well as residential customers and had extended to a ~5 km loop. The roots of DC go back to the 19th century. It was initially introduced as a scheme to distribute clean, cool air to houses through underground pipes. The first known DC system began operations at Denver's Colorado Automatic Refrigerator Company in late 1889. In the 1930s, large DC systems were created for Rockefeller Center in New York City and for the U.S. Capitol in Washington, DC.

It is believed that district energy in Canada began in London, Ontario in 1880. The London system was built in the form of a group of systems serving the university, hospital, and government complexes. The University of Toronto is known to have developed a DH system in 1911 that served the needs of the university. The first commercial DH system in Canada was established in 1924, in the city of Winnipeg's commercial core. Canada boasts the site of one of the northernmost DESs in North America: Fort McPherson, located in the Northwest Territories. The Canadian District Energy Association (CDEA) was created in 1993 in recognition of the fact that the emerging Canadian district energy industry needed to create a common voice to promote DHC applications. It aims to exchange and share information and experience for its stakeholders. It has also been instrumental in helping provide a forum for the exchange of ideas and information, and in identifying and addressing key technical and policy issues to advance the use of district energy in Canada.[5]

TECHNICAL ASPECTS OF DISTRICT HEATING AND COOLING

District heating and cooling is the distribution of heating (hot water, steam) and cooling (cold water) energy transfer mediums from a central energy production source to meet the diverse thermal energy needs of residential, commercial, and industrial users. Thermal energy needs or demands include space heating and cooling systems for maintaining human comfort, domestic hot water requirements, manufacturing-plant process heating and cooling system requirements, etc. In many of the systems that have been established around the world, both district heating and district cooling have not been provided. In Europe, for example, where moderate summer temperatures prevail, most DESs provide heating capability only. DC has only recently become more widespread, with the most prevalent application being in North America, where summer temperatures can reach extremes of 30°C–40°C over extended periods.

To implement a DH, DC, or DHC in a community, one should weigh several factors to conduct a feasibility study for determining whether a DH, DC, or DHC system is suitable for application. Essential factors include energetic, environmental, economic, and social criteria; operating conditions; fuel availability; efficiency considerations; local benefits; viability of competing systems; local climatic conditions; user characteristics such as load density; total load requirements; characteristics of the heating and cooling systems currently in place; developer's perspectives; and local utility considerations.

As shown in Fig. 1, a basic DHC consists of three subsystems—namely, energy generation, energy transmission, and energy use. These subsystems are described below:

1. *Energy generation*. In this section, steam or hot water (in the case of DH) and chilled or cold water (in the case of DC) are produced.
2. *Energy transmission*. In this part, the thermal energy medium (steam or water) is distributed via pipelines from the production sources to the network of users.
3. *Energy use*. In this section, energy (either heat or cold) is consumed for either heating or cooling purposes in the facilities. A combination of residential, public, commercial, and industrial users may be involved with varying uses of the thermal energy, including space heating and cooling, domestic water heating, and plant process heating and cooling.

Note that a DH and/or DC system differs fundamentally from a conventional system in that in the latter system, thermal energy is produced and distributed at the location of use. Examples of conventional systems include residential heating and cooling with, respectively, furnaces and air conditioners; electric heating of offices; package boilers/chillers providing heating/cooling of apartment complexes; and a dedicated boiler plant providing heat to an industrial facility.

DISTRICT HEATING AND COOLING AND ENVIRONMENTAL IMPACT

Problems with energy supply and use are related not only to global warming, but also to such environmental concerns as air pollution, acid precipitation, ozone depletion, forest destruction, and emission of radioactive substances. These issues must be taken into consideration simultaneously if humanity is to achieve a bright energy future with minimal environmental impact. Much evidence exists to suggest that the future will be negatively impacted if humans keep degrading the environment.

One solution to both energy and environmental problems is to enhance much more use of DHC applications.

Numerous fuels are used at DHC plants, including various grades of oil and coal, natural gas, refuse, and other biofuels (e.g., wood chips, peat, and straw). The combustion of such fuels may produce environmentally hazardous products of combustion; thus, flue-gas cleaning devices and other emission reduction measures are often incorporated. Some measures are usually required under increasingly strict legislation before approval to operate a facility is granted. Examples of pollution control equipment used at DHC plants include acid–gas scrubbers. These systems typically utilize hydrated lime to react with the moisture, SO, and other acid gases in the flue gases discharged from the combustion system. With such systems, the lime–acid, gas–water vapor reaction products are efficiently collected by electrostatic precipitators as particulate matter. Bag filters are also utilized in many applications to capture the particulate matter as well as the acid–gas scrubbing reaction products. Conventional oil-/gas-fired boilers utilizing low NO_x, burners to reduce NO_x emissions dramatically are also becoming more common. Flue-gas recirculation to reduce NO_x emissions has also proved to be effective. Other emission control or reduction techniques can be introduced with DHC systems, including optimization of combustion efficiency (i.e., reduced CO_2, CO, and hydrocarbon emissions) through the use of modern computerized combustion control systems, and utilization of higher-quality, lower-emission-producing fuels. With the above, it is apparent that heating and cooling systems that minimize the quantity of fuel and electrical power required to meet users' needs will result in reduced impact on the environment.

In addition, DHC systems that comprise several types of thermal energy generation plants can optimize plant and system efficiency by utilizing, whenever possible, the thermal energy sources with the highest energy conversion efficiencies for base and other partial load conditions. Then the sources with the poorer conversion efficiencies can be utilized only to meet peak loads. Essentially, improved efficiency means use of less fuel for the same amount of energy produced, which in turn results in the conservation of fossil fuels, reduced emissions of pollutants, improved air quality, and reduced use of CFC refrigerants (if any) in DC applications.

District heating and cooling systems are well suited to combine with electric power production facilities as cogeneration plants. The amalgamation of these two energy production/utilization schemes results in a substantial improvement in overall energy conversion efficiency, because DH systems can effectively utilize the otherwise-wasted heat associated with the electric power production process. A district system meeting much or all of its load requirements with waste heat from power generation facilities will have a positive environmental impact, as fuel consumption within the community is reduced considerably. Conservation of fossil fuels and a reduction of combustion-related emissions are resultant direct benefits of such a DHC system.

The centralized nature of DHC energy production plants results in a reduced number of emissions sources in a community. This introduces the potential for several direct benefits.

The higher operating efficiency afforded to larger, well-maintained facilities translates directly to reduced fuel consumption, which in turn results in conservation of fossil fuels and reduced emissions. Higher operating efficiency of the combustion process (where parameters such as temperature, combustion air and fuel input levels, and residence time are closely monitored) also impacts emission production in that the concentration of certain pollutants produced—particularly CO_2 and NO_x—is reduced.

Furthermore, measures to increase energy efficiency can reduce environmental impact by reducing energy losses. From an exergy viewpoint, such activities lead to increased exergy efficiency and reduced exergy losses (both waste exergy emissions and internal exergy consumption).

A deeper understanding of the relations between exergy and the environment may reveal the underlying fundamental patterns and forces affecting changes in the environment, and help researchers deal better with environmental damage.

The second law of thermodynamics is instrumental in providing insights into environmental impact. The most appropriate link between the second law and environmental impact has been suggested to be exergy, in part because it is a measure of the departure of the state of a system from that of the environment. The magnitude of the exergy of a system depends on the states of both the system and the environment. This departure is zero only when the system is in equilibrium with its environment.

To achieve the energetic, economic, and environmental benefits that DHCs offer, the following integrated set of activities should be carried out[6]:

- *Research and development*. Research and development priorities should be set in close consultation with industry to reflect its needs. Most research is conducted through cost-shared agreements and falls within the short-to-medium term. Partners in these activities should include a variety of stakeholders in the energy industry, such as private-sector firms, utilities across the country, provincial governments, and other federal departments.
- *Technology assessment*. Appropriate technical data should be gathered in the lab and through field trials on factors such as cost benefit, reliability, environmental impact, safety, and opportunities for improvement. These data should also assist the preparation of technology status overviews and strategic plans for further research and development.
- *Standards development*. The development of technical and safety standards is needed to encourage the acceptance of proven technologies in the marketplace. Standards development should be conducted in cooperation with national and international standards-writing organizations, as well as with other national and provincial regulatory bodies.
- *Technology transfer*. Research and development results should be transferred through sponsorship of technical workshops, seminars, and conferences, as well as through the development of training manuals and design tools, Web tools, and the publication of technical reports.

Such activities will also encourage potential users to consider the benefits of adopting DHC applications, using renewable energy resources. In support of developing near-term markets, a key technology transfer area is to accelerate the use of cogeneration and DHC applications, particularly for better efficiency, cost effectiveness, and environmental considerations.

DISTRICT HEATING AND COOLING AND SUSTAINABLE DEVELOPMENT

Sustainable development requires a sustainable supply of clean and affordable energy resources that do not cause negative societal impacts.[7] Supplies of such energy resources as fossil fuels and uranium are finite. Green energy resources (e.g., solar and wind) are generally considered to be renewable and, therefore, sustainable over the relatively long term.

Sustainability often leads local and national authorities to incorporate environmental considerations into energy planning. The need to satisfy basic human needs and aspirations, combined with increasing world population, will make the need for successful implementation of sustainable development increasingly apparent. Various criteria that are essential to achieving sustainable development in a society follow:

- Information about and public awareness of the benefits of sustainability investments

- Environmental education and training
- Appropriate energy and exergy strategies
- The availability of renewable energy sources and cleaner technologies
- A reasonable supply of financing
- Monitoring and evaluation tools

The key point here is to use renewable energy resources in DHC systems. As is known, not all renewable energy resources are inherently clean, in that they cause no burden on the environment in terms of waste emissions, resource extraction, or other environmental disruptions. Nevertheless, the use of DHC almost certainly can provide a cleaner and more sustainable energy system than increased controls on conventional energy systems.

To seize the opportunities, it is essential to establish a DHC market and gradually build up the experience with cutting-edge technologies. The barriers and constraints to the diffusion of DHC use should be removed. The legal, administrative, and financing infrastructure should be established to facilitate planning and application of geothermal energy projects. Government could and should play a useful role in promoting geothermal energy technologies through funding and incentives to encourage research and development, as well as commercialization and implementation in both urban and rural areas.

Environmental concerns are significantly linked to sustainable development. Activities that continually degrade the environment are not sustainable. The cumulative impact on the environment of such activities often leads over time to a variety of health, ecological, and other problems. Clearly, a strong relationship exists between efficiency and environmental impact, because for the same services or products, less resource utilization and pollution are normally associated with increased efficiency.[8]

Improved energy efficiency leads to reduced energy losses. Most efficiency improvements produce direct environmental benefits in two ways. First, operating energy input requirements are reduced per unit output, and the pollutants generated are reduced correspondingly. Second, consideration of the entire life cycle for energy resources and technologies suggests that improved efficiency reduces environmental impact during most stages of the life cycle.

In recent years, the increased acknowledgment of humankind's interdependence with the environment has been embraced in the concept of sustainable development. With energy constituting a basic necessity for maintaining and improving standards of living throughout the world, the widespread use of fossil fuels may have impacted the planet in ways far more significant than first thought. In addition to the manageable impacts of mining and drilling for fossil fuels, and discharging wastes from processing and refining operations, the "greenhouse" gases created by burning these fuels is regarded as a major contributor to a global-warming threat. Global warming and large-scale climate change have implications for food-chain disruption, flooding, and severe weather events.

Use of renewable energy sources in DHC systems with cogeneration can help reduce environmental damage and achieve sustainability.

Sustainable development requires not just that sustainable energy resources be used, but also that the resources be used efficiently. The authors and others feel that exergy methods can be used to evaluate and improve efficiency, and thus to improve sustainability. Because energy can never be "lost," as it is conserved according to the first law of thermodynamics, whereas exergy can be lost due to internal irreversibilities, this study suggests that exergy losses that represent potential not used, particularly from the use of nonrenewable energy forms, should be minimized when striving for sustainable development. In the next section, the authors discuss the exergetic aspects of thermal systems and present an efficiency analysis for performance improvement.

Furthermore, this study shows that some environmental effects associated with emissions and resource depletion can be expressed based on physical principles in terms of an exergy-based indicator. It may be possible to generalize this indicator to cover a comprehensive range of environmental effects, and research in line with that objective is ongoing.

Although this work discusses the benefits of using thermodynamic principles—especially exergy—to assess the sustainability and environmental impact of energy systems, this area of work is relatively new. Further research is needed to gain a better understanding of the potential role of exergy in such a comprehensive perspective. This includes the need for research to (1) better define the role of exergy in environmental impact and design; (2) identify how exergy can be better used as an indicator of potential environmental impact; and (3) develop holistic exergy-based methods that simultaneously account for technical, economic, environmental, and other factors.

PERFORMANCE EVALUATION

From the thermodynamics point of view, exergy is defined as the maximum amount of work that can be produced by a system or a flow of matter or energy as it comes to equilibrium with a reference environment. Exergy is a measure of the potential of the system or flow to cause change, as a consequence of not being completely in stable equilibrium relative to the reference environment. Unlike energy, exergy is not subject to a conservation law (except for ideal, or reversible, processes). Rather, exergy is consumed or destroyed, due to irreversibilities in any real process. The exergy consumption during a process is proportional to the entropy created due to irreversibilities associated with the process.

Exergy analysis is a technique that uses the conservation of mass and conservation of energy principles together with the second law of thermodynamics for the analysis, design, and improvement of DHC systems, as well as others. It is also useful for improving the efficiency of energy-resource use, for it quantifies the locations, types, and magnitudes of waste and loss. In general, more meaningful efficiencies are evaluated with exergy analysis rather than energy analysis, because exergy efficiencies are always a measure of how nearly the efficiency of a process approaches the ideal. Therefore, exergy analysis identifies accurately the margin available to design more efficient energy systems by reducing inefficiencies. We can suggest that thermodynamic performance is best evaluated using exergy analysis because it provides more insights and is more useful in efficiency-improvement efforts than energy analysis. For exergy analysis, the characteristics of a reference environment must be specified. This is commonly done by specifying the temperature, pressure, and chemical composition of the reference environment. The results of exergy analyses, consequently, are relative to the specified reference environment, which in most applications is modeled after the actual local environment. The exergy of a system is zero when it is in equilibrium with the reference environment.

In exergy analysis, the temperatures at different points in the system are important, and the equivalent temperature T_{equiv} between the supply (Eq. 1) and return (Eq. 2) temperatures can be written as

$$T_{\text{equiv}} = \frac{h_1 - h_2}{s_1 - s_2} \tag{1}$$

where h and s denote specific enthalpy and specific entropy, respectively.

Here in the cogeneration section, the electricity production rate $\dot{W}$ can be expressed for a cogeneration-based system using electric chillers as a function of the product-heat generation rate $\dot{Q}_{\text{H}}$ as

$$\dot{W} = \left(\frac{\eta_{\text{elec}}^{\text{CHP}}}{\eta_{\text{heat}}^{\text{CHP}}}\right)\dot{Q}_{\text{H}} \tag{2}$$

and for a cogeneration-based system using absorption chillers as a function of the product-heat generation rates, $\dot{Q}_{\text{H}}$ and $\dot{Q}_{\text{gen}}$, as

$$\dot{W} = \left(\frac{\eta_{\text{elec}}^{\text{CHP}}}{\eta_{\text{heat}}^{\text{CHP}}}\right)(\dot{Q}_{\text{H}} + \dot{Q}_{\text{gen}}) \tag{3}$$

where $\eta_{\text{elec}}^{\text{CHP}}$ and $\eta_{\text{heat}}^{\text{CHP}}$ denote, respectively, the electrical and heat efficiencies of the cogeneration, or CHP, plant.

The total energy efficiency can be written for the cogeneration plant using electric chillers as

$$\eta^{\text{CHP}} = \frac{\dot{W} + \dot{Q}_{\text{H}}}{\dot{E}_{\text{f}}} \tag{4}$$

and for the cogeneration plant using absorption chillers as

$$\eta^{\text{CHP}} = \frac{\dot{W} + \dot{Q}_{\text{H}} + \dot{Q}_{\text{gen}}}{\dot{E}_{\text{f}}} \tag{5}$$

where $\dot{E}_{\text{f}}$ denotes the fuel energy input rate. The corresponding total exergy efficiency can be expressed for the cogeneration plant using electric chillers as

$$\psi^{\text{CHP}} = \frac{\dot{W} + \tau_{Q_{\text{H}}}\dot{Q}_{\text{H}}}{R\dot{E}_{\text{f}}} \tag{6}$$

and for the cogeneration plant using absorption chillers as

$$\psi^{\text{CHP}} = \frac{\dot{W} + \tau_{Q_{\text{H}}}\dot{Q}_{\text{H}} + \tau_{Q_{\text{gen}}}\dot{Q}_{\text{gen}}}{R\dot{E}_{\text{f}}} \tag{7}$$

where $\tau_{Q_{\text{H}}}$ and $\tau_{Q_{\text{gen}}}$ are the exergetic temperature factors for $\dot{Q}_{\text{H}}$ and $\dot{Q}_{\text{gen}}$, respectively.

For heat transfer at a temperature T, the exergetic temperature factor can be written as

$$\tau \equiv 1 - \frac{T_0}{T} \tag{8}$$

The fuel exergy flow rate, which is all chemical exergy, is evaluated as $R\dot{E}_{\text{f}}$ here, where R and $\dot{E}_{\text{f}}$ denote, respectively, the energy grade function and energy flow rate of the fuel.

In the chilling process, the exergy efficiency can be written for the chilling operation using electric chillers as

$$\psi_{\text{ch}} = \frac{-\tau_{Q_{\text{C}}}\dot{Q}_{\text{C}}}{\dot{W}_{\text{ch}}} \tag{9}$$

and using absorption chillers as

$$\psi_{\text{ch}} = \frac{-\tau_{Q_{\text{C}}}\dot{Q}_{\text{C}}}{\tau_{Q_{\text{gen}}}\dot{Q}_{\text{gen}}} \tag{10}$$

In this part, we deal with DHC. District heating utilizes hot-water supply and warm-water return pipes, whereas DC utilizes cold-water supply and cool-water return pipes. The pipes are assumed to be perfectly insulated so that heat loss or infiltration during fluid transport can be minimized. Hence, the energy efficiencies of the DHC portions of the system are both 100%. The exergy efficiency can be evaluated for DH as

$$\psi_{\text{DH}} = \frac{\tau_{Q_{\text{H}}^{\text{u}}}\dot{Q}_{\text{H}}^{\text{u}}}{\tau_{Q_{\text{H}}}\dot{Q}_{\text{H}}} \tag{11}$$

and for DC as

$$\psi_{\text{DC}} = \frac{-\tau_{Q_{\text{C}}^{\text{u}}}\dot{Q}_{\text{C}}^{\text{u}}}{-\tau_{Q_{\text{C}}}\dot{Q}_{\text{C}}} \tag{12}$$

Heat loss and infiltration for the user heating and cooling subsystems are assumed to be negligible, so that

Dem–Dry

their energy efficiencies are assumed to be 100%. The exergy efficiency can be expressed for the user-heating subsystem as

$$\psi_{\mathrm{UH}} = \frac{\tau_{Q_{\mathrm{H}}^{\mathrm{u,s}}} \dot{Q}_{\mathrm{H}}^{\mathrm{u,s}} + \tau_{Q_{\mathrm{H}}^{\mathrm{u,w}}} \dot{Q}_{\mathrm{H}}^{\mathrm{u,w}}}{\tau_{Q_{\mathrm{H}}^{\mathrm{u}}} \dot{Q}_{\mathrm{H}}^{\mathrm{u}}} \quad (13)$$

and for the user-cooling subsystem as

$$\psi_{\mathrm{UC}} = \frac{-\tau_{Q_{\mathrm{C}}^{\mathrm{u,r}}} \dot{Q}_{\mathrm{C}}^{\mathrm{u,r}}}{-\tau_{Q_{\mathrm{C}}^{\mathrm{u}}} \dot{Q}_{\mathrm{C}}^{\mathrm{u}}} \quad (14)$$

The left and right terms in the numerator of Eq. 13 represent the thermal exergy supply rates for space and water heating, respectively.

For the overall process, because three different products (electricity, heat, and cooling) are generated, application of the term energy efficiency here is prone to be misleading, in part for the same reason that the term energy efficiency is misleading for a chiller. Here, an overall-system "figure of merit" f_{sys} is used, calculated as follows:

$$f_{\mathrm{sys}} = \frac{\dot{W}_{\mathrm{net}} + \dot{Q}_{\mathrm{H}}^{\mathrm{u,s}} + \dot{Q}_{\mathrm{H}}^{\mathrm{u,w}} + \dot{Q}_{\mathrm{C}}^{\mathrm{u,r}}}{\dot{E}_{\mathrm{f}}} \quad (15)$$

The corresponding exergy-based measure of efficiency is simply an exergy efficiency and is evaluated as

$$\psi_{\mathrm{sys}} = \frac{\dot{W}_{\mathrm{net}} + \tau_{Q_{\mathrm{H}}^{\mathrm{u,s}}} \dot{Q}_{\mathrm{H}}^{\mathrm{u,s}} + \tau_{Q_{\mathrm{H}}^{\mathrm{u,w}}} \dot{Q}_{\mathrm{H}}^{\mathrm{u,w}} - \tau_{Q_{\mathrm{C}}^{\mathrm{u,r}}} \dot{Q}_{\mathrm{C}}^{\mathrm{u,r}}}{R\dot{E}_{\mathrm{f}}} \quad (16)$$

For further details about the energy and exergy analysis of the systems.[9,10]

CASE STUDIES

Here, we present an efficiency analysis, accounting for both energy and exergy considerations, for two case studies—namely, a cogeneration-based DES and a geothermal district heating system (GDHS).

Case Study I

The case considered here for analysis is a major cogeneration-based DHC project in downtown Edmonton, Alberta,[11,12] having (1) an initial supply capacity of 230 MW (thermal) for heating and 100 MW (thermal) for cooling; (2) the capacity to displace about 15 MW of electrical power used for electric chillers through DC; and (3) the potential to increase the efficiency of the Rossdale power plant that would cogenerate to provide the steam for DHC from about 30 to 70%, respectively. The design includes the potential to expand the supply capacity for heating to about 400 MW (thermal). The design incorporated central chillers and a DC network. Screw chillers were to be used originally and absorption chillers in the future. Central chillers are often favored because (1) the seasonal efficiency of the chillers can increase due to the ability to operate at peak efficiency more often in a large central plant; and (2) lower chiller condenser temperatures (e.g., 20°C) can be used if cooling water from the environment was available to the central plant, relative to the condenser temperatures of approximately 35°C needed for air-cooled building chillers. These two effects can lead to large central chillers having almost double the efficiencies of distributed small chillers.

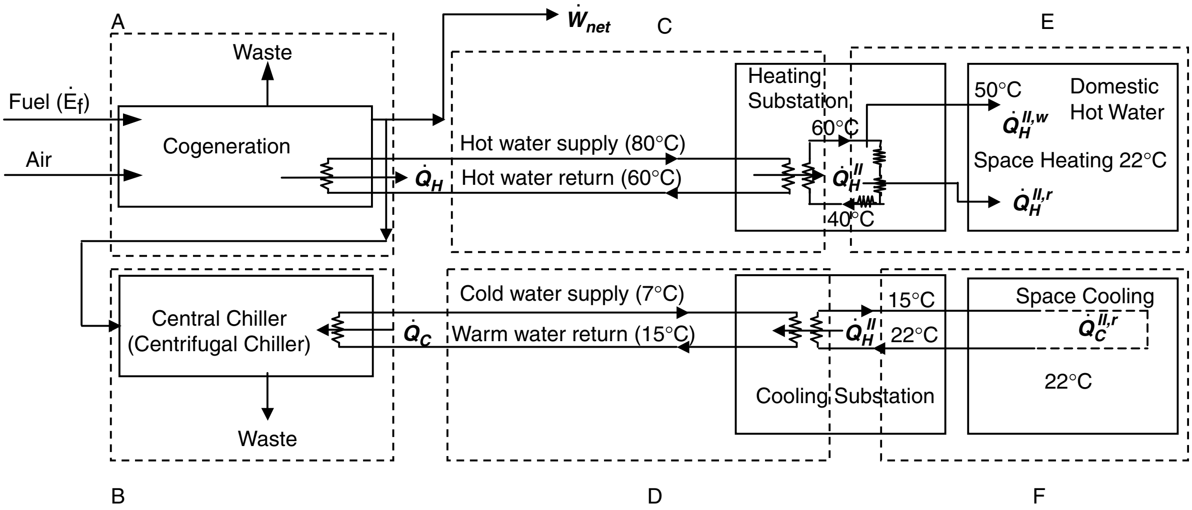


Fig. 3 Simplified diagram of the cogeneration-based district energy system (DES) in the Edmonton Power. The system, which uses electric chillers, is divided into six subsections within three categories. On the left are production processes, including cogeneration of electricity and heat (A) and chilling (B). In the middle are district-energy transport processes, including district heating (DH) (C) and DC (D). On the right are end-user processes, including user heating (E) and user cooling (F).
Source: Adapted from Refs. [9,10].

Table 2 Monthly heating and cooling load breakdown (in %) in the design area of Edmonton, Alberta

	Period 1 (Winter)								Period 2 (Summer)					
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	Total	May	June	July	Aug	Sep	Total
Heating	6.90	12.73	16.83	18.67	14.05	12.95	7.34	89.46	2.39	1.56	1.34	1.92	3.33	10.54
Cooling	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.62	22.06	32.00	26.80	8.52	100

Source: Adapted from Refs. [9–11].

There are two main stages in this case study, taken from.[9,10] First, the design for cogeneration-based DHC[11,12] is evaluated thermodynamically. Then the design is modified by replacing the electric centrifugal chillers with heat-driven absorption chillers (first single- and then double-effect types) and reevaluating it thermodynamically.

The cogeneration-based DES considered here (see Fig. 3) includes a cogeneration plant for heat and electricity, and a central electric chiller that produces a chilled fluid. Hot water was produced to satisfy all heating requirements of the users, at a temperature and pressure of 120°C and 2 bar, respectively. The heat was distributed to the users via heat exchangers, DH grids, and user heat-exchanger substations. A portion of the cogenerated electricity was used to drive a central centrifugal chiller, and the remaining electricity was used for other purposes (export, driving other electrical devices, etc.). The central chiller produces cold water at 7°C, which was distributed to users via DC grids.

For the cogeneration-based DES using absorption chillers, the design was modified by replacing the electric chiller with single-effect absorption chillers. Hot water was produced at 120°C and 2 bar to satisfy all heating requirements of the users and to drive the central absorption chillers. A small portion of the cogenerated electricity was used to drive the absorption solution and refrigeration pumps, and the remaining electricity was used for purposes other than space cooling. Then this cogeneration-based DES was further modified by replacing the electric centrifugal chillers with double-effect absorption chillers. The system was similar to the cogeneration-based DES using single-effect absorption chillers, except that higher-quality heat (170°C and 8 bar) was produced to drive the double-effect absorption chillers.

For the analysis, the year was divided into two seasonal periods (see Table 2). Period 1 (October to April) has an environmental temperature of 0°C and was considered to be a winter period with only a heating demand. Period 2 (May to September) has an environmental temperature of 30°C and was considered to be a summer period with a cooling demand and a small demand for water heating. The small variations in plant efficiency that occur with changes in environmental temperature are neglected here.

The overall energy efficiency of the proposed cogeneration plant was 85%; the electrical efficiency (i.e., the efficiency of producing electricity via cogeneration) was 25%; and the heat production efficiency was 60%. Also, the total heating requirement of the buildings in the design region was $\dot{Q}_H = 1040$ GWh/yr for space and water heating, and the cooling requirement was $\dot{Q}_C = 202$ GWh/yr for space cooling. The total fuel energy input rate can be evaluated for cogeneration plant using electric chillers as $\dot{E}_f = 1040/0.6 = 1733$ GWh/yr. Because 33 GWh/yr of this cooling was provided through free cooling, the cooling requirement of the chilling plant was 169 GWh/yr.[11] The COP of the single-effect absorption chiller used here was taken to be 0.67, a typical representative value. Therefore, the annual heat required to drive the single-effect absorption machine was $\dot{Q}_{gen} = 169/0.67 = 252$ GWh/yr. Thus, the total fuel energy input rate to the cogeneration plant can be evaluated as $\dot{E}_f = (1040 + 252)/0.6 = 2153$ GWh/yr.[9,10]

As mentioned above, steam was required at higher temperatures and pressures to drive the double-effect absorption chillers, and more electricity was curtailed as higher quality of heat or more heat was produced. The overall energy efficiency of the proposed cogeneration plant was unchanged (85%) in Period 2. Only the electrical and heat efficiencies changed due to more heat being produced in this period, when the absorption chiller was in operation. Thus, the electrical efficiency (i.e., the efficiency of producing electricity via cogeneration) was 25 and 21%, respectively, in Periods 1 and 2, respectively, and the heat production efficiency was 60 and 64%, respectively, in Periods 1 and 2, respectively. The COP of the double-effect absorption chiller used here was taken to be 1.2, a typical representative value. Therefore, the annual heat required to drive the double-effect absorption machine was $\dot{Q}_{gen} = 169/1.2 = 141$ GWh/yr. The total fuel energy input rate to the cogeneration plant can be evaluated as the sum of the fuel energy input rate to the plant in two periods. Thus, $\dot{E}_f = 1942$ GWh/yr.[9,10]

The average supply and return temperatures, respectively, were taken as 80 and 60°C for DH, and 7 and 15°C for DC. The supply and return temperatures, respectively, were taken as 60 and 40°C for the user-heating substation, and 15 and 22°C for the user-cooling substation. Furthermore, the user room temperature was considered constant throughout the year at 22°C. For DH the equivalent temperature was 70°C for the supply system and 50°C for the user substation, whereas for DC the equivalent temperature was 11°C for the supply system and 19°C for the user substation.

Table 2 shows that 89.46 and 10.54% of the total annual heat loads occur in Periods 1 and 2, respectively. Because there was assumed to be no space heating demand in Period 2, the 10.54% quantity was taken to be the heat needs for water heating (which was assumed to be constant throughout the year). Table 2 also presents the space cooling breakdown in Period 2. Annual energy transfer rates for the cogeneration-based DES are shown in Table 3, with details distinguished where appropriate for the three chiller options considered. The data in Table 3 are used to calculate exergy efficiencies for the systems for each period and for the year.

Edmonton Power had annual free cooling of 33 GWh/yr; the cooling requirement of the chilling plant was 169 GWh/yr. The COP of the centrifugal chiller in the design was 4.5. Thus, the annual electricity supply rate to the chiller was $\dot{W}_{ch} = 169/4.5 = 38$ GWh/yr. For the chilling operation, including free cooling and electrical cooling, COP = (169 + 33)/38 = 5.32. The net electricity output ($\dot{W}_{net}$) of the combined cogeneration/chiller portion of the system was 433 − 38 = 395 GWh/yr, where the electrical generation rate of the cogeneration plant was 433 GWh/yr. Similarly, for the chilling operation, including free cooling and single-effect absorption cooling, the coefficient of performance was COP = 202/252 = 0.80,

Table 3 Annual Energy Transfer Rates (in GWh/yr) for the Cogeneration-Based DHC System in Edmonton, Alberta

Type of Energy	Period 1, $T_o = 0$°C	Period 2, $T_o = 30$°C
District heating, $\dot{Q}_H$	0.8946 × 1040 = 930	0.1054 × 1040 = 110
Water heating, $\dot{Q}_H^{u,w}$	(22 GWh/yr/mo.) × 7 mo. = 154	0.1054 × 1040 = 110 (or 22 GWh/yr/mo.)
Space heating, $\dot{Q}_H^{u,s}$	930–154 = 776	0
Space cooling, $\dot{Q}_C$	0	1.00 × 202 = 202
Electric chiller case		
Total electricity, $\dot{W}$	0.8946 × 433 = 388	0.1054 × 433 = 45.6
Input energy, $\dot{E}_f$	0.8946 × 1733 = 1551	0.1054 × 1733 = 183
Single-effect absorption chiller case		
Heat to drive absorption chiller, $\dot{Q}_{gen}$	0	1.00 × 252 = 252
Total electricity, $\dot{W}$	0.8946 × 433 = 388	25/60 (110 + 252) = 151
Input energy, $\dot{E}_f$	0.8946 × 1733 = 1551	(110 + 252)/0.6 = 603
Double-effect absorption chiller case		
Heat to drive absorption chiller, $\dot{Q}_{gen}$	0	1.00 × 141 = 141
Total electricity, $\dot{W}$	0.8946 × 433 = 388	21/64 × (110 + 141) = 82
Input energy, $\dot{E}_f$	0.8946 × 1733 = 1551	(110 + 141)/0.64 = 391

Source: Adapted from Refs. [9,10].

Dem–Dry

Table 4 System and subsystem efficiencies for the cogeneration-based DES for several types of chillers

System	Efficiency (%) Energy (η) Centrifugal chiller	Efficiency (%) Energy (η) 1-Stage absorption chiller	Efficiency (%) Energy (η) 2-Stage absorption chiller	Efficiency (%) Exergy (ψ) Centrifugal chiller	Efficiency (%) Exergy (ψ) 1-Stage absorption chiller	Efficiency (%) Exergy (ψ) 2-Stage absorption chiller
Individual Subsystems						
Cogeneration	85	85	85	37	37	37
Chilling	450[a]	67[a]	120[a]	36	23	30
District heating (DH)	100	100	100	74	74	74
District cooling (DC)	100	100	100	58	58	58
User heating (UH)	100	100	100	54	54	54
User cooling (UC)	100	100	100	69	69	69
Combination subsystems[b]						
Cogeneration + chilling	94	83	88	35	35	35
District energy (DE)	100	100	100	73	73	73
User energy (UE)	100	100	100	53	53	53
Cogeneration + DH	85	85	85	34	35	34
Cogeneration + DH + UH	85	85	85	30	31	31
Chilling + DC	532[a]	80[a]	143[a]	21	14	18
Chilling + DC + UC	532[a]	80[a]	143[a]	14	9	12
DH + UH	100	100	100	40	40	40
DC + UC	100	100	100	41	41	41
Cogeneration + chilling + DE	94	83	88	32	32	32
DE + UE	100	100	100	40	40	40
Overall process	94	83	88	28	29	29

[a] These are coefficient of performance (*COP*) values when divided by 100.
[b] DE = DH + DC and UE = UH + UC.
Source: Adapted from Refs. [9,10].

and for double-effect absorption cooling, it was COP = 202/141 = 1.43. It should be noted that the work required to drive the solution and refrigeration pumps was very small relative to the heat input to the absorption chiller (often less than 0.1%); this work was thus neglected here.

Table 4 lists the energy and exergy efficiencies evaluated for the individual subsystems, several subsystems comprised of selected combinations of the individual subsystems, and the overall system for cogeneration-based DES using electric chillers, single-effect absorption chillers, and double-effect absorption chillers. Overall energy efficiencies are seen to vary for the three system alternatives considered, from 83 to 94%, respectively, and exergy efficiencies vary from 28 to 29%, respectively. Table 4 demonstrates that energy efficiencies do not provide meaningful and comparable results relative to exergy efficiencies when the energy products are in different forms. The energy efficiency of the overall process using electric chillers, for example, is 94%, which could lead one to believe that the system is very efficient. The exergy efficiency of the overall process, however, is 28%, indicating that the process is far from ideal thermodynamically. The exergy efficiency is much lower than energy efficiency because the heat is being produced at a temperature (120°C) much higher than the temperatures actually needed (22°C for space heating and 40°C for water heating). The low exergy efficiency of the chillers is largely responsible for the low exergy efficiency for the overall process. The exergy-based efficiencies in Table 4 are generally lower than the energy-based ones because the energy efficiencies utilize energy quantities that are in different forms, whereas the exergy efficiencies provide more meaningful and useful results by evaluating the performance and behavior of the systems using electrical equivalents for all energy forms. The results for cogeneration-based DESs using absorption chillers (single-effect and double-effect absorption chillers) and those using electric chillers are, in general, found to be similar.[11,12]

For cogeneration-based district energy, in which electricity, heating, and cooling are produced simultaneously, exergy analysis provides important insights into the performance and efficiency for an overall system and its separate components. This thermodynamic analysis technique provides more meaningful efficiencies than energy analysis, and pinpoints the locations and causes of inefficiencies more accurately. The present results indicate that the complex array of energy forms involved in cogeneration-based DESs make them difficult to assess and compare thermodynamically without exergy analysis. This difficulty is attributable primarily to the different nature and quality of the three product energy forms: electricity, heat, and cooling. The results are expected to aid designers of such systems in development and optimization activities, and in selecting the proper type of system for different applications and situations.

Case Study II

Geothermal district heating has been given increasing attention in many countries during the past decade, and many successful geothermal district heating projects have been reported. For district heating to become a serious alternative to existing or future individual heating and/or cooling systems, it must provide significant benefits to both the community in which it is operated and the consumers who purchase energy from the system. Further, it must provide major societal benefits if federal, state, or local governments are to offer the financial and/or institutional support that is required for successful development.[13]

The case study here is the Izmir–Balcova GDHS, which is one example of the high-temperature district heating applications in Turkey. The Balcova region is about 7 km from the Centrum of the Izmir province, located in western Turkey, and is endowed with considerably rich geothermal resources. The Izmir–Balcova geothermal field (IBGF) covers a total area of about 3.5 km^2 with an average thickness of the aquifer horizon of 150 m. In the district heating system investigated, there are two systems—namely, the Izmir–Balcova GDHS and the Izmir–Narlidere GDHS. The design heating capacity of the Izmir–Balcova GDHS is equivalent to 7500 residences. The INGDHS was designed for 1500 residence equivalence but has a sufficient infrastructure to allow capacity growth to 5000 residence equivalence. The outdoor and indoor design temperatures for the two systems are 0 and 22°C, respectively. Fig. 4 illustrates a schematic of the IBGF, where the Izmir–Balcova GDHS, the Izmir–Narlidere GDHS, and hotels and official buildings heated by geothermal energy were included. The Izmir–Balcova GDHS consists mainly of three cycles, such as (a) energy production cycle (geothermal well loop and geothermal heating center loop), (b) energy distribution cycle (district heating distribution network), and (c) energy consumption cycle (building substations). As of the end of 2001, there are 14 wells ranging in depth from 48 to 1100 m in the IBGF. Of those, seven and six wells are production and reinjection wells, respectively, while one well is out of operation. The wellhead temperatures of the production wells vary from 95 to 140°C, with an average value of 118°C, and the volumetric flow rates of the wells range from 30 to 150 m^3/h. Geothermal fluid, collected from the seven production wells at an average wellhead temperature of 118°C, is pumped to a mixing chamber, where it is mixed with the reinjection fluid at an average temperature of 60°C–62°C, cooling the mixture to 98°C–99°C. Then this geothermal fluid is sent to two primary-plate-type heat exchangers and cooled to about 60°C–62°C as its heat is transferred to the secondary fluid. The geothermal fluid whose heat is taken at the geothermal center is reinjected into the reinjection wells, while the secondary fluid (clean hot water) is transferred to the heating circulation water of

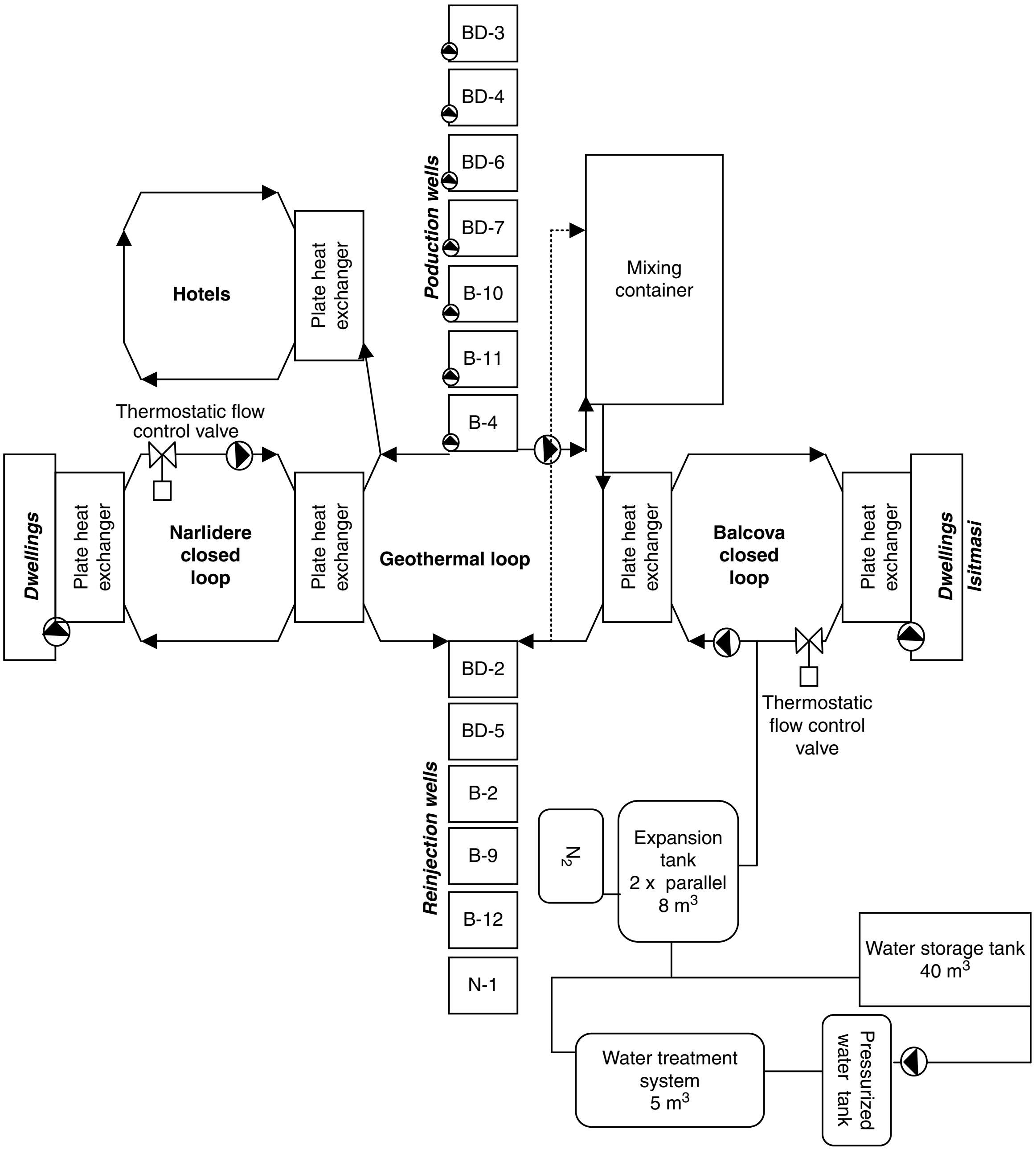


Fig. 4 A schematic of the Izmir–Balcova–Narlıdere geothermal district heating system.

the building by the heat exchangers of the substations. The average conversion temperatures obtained during the operation of the IBGDHS are, on average, 80°C/57°C for the district heating distribution network and 65°C/45°C for the building circuit. By using the control valves for flow rate and temperature at the building substations, the needed amount of water is sent to each housing unit, and the heat balance of the system is achieved.[14]

In the following paragraphs, we give main relations for mass, energy, and exergy flows, along with the energy and exergy efficiencies for the Izmir–Balcova GDHS.[15]

The mass balance equation is written as follows:

$$\sum_{i=1}^{n} \dot{m}_{\mathrm{w,tot}} - \dot{m}_{\mathrm{r}} - \dot{m}_{\mathrm{d}} = 0 \tag{17}$$

where $\dot{m}_{\mathrm{w,tot}}$ is the total mass flow rate at the wellhead, $\dot{m}_{\mathrm{r}}$ is the flow rate of the reinjected thermal water, and $\dot{m}_{\mathrm{d}}$ is the mass flow rate of the natural direct discharge.

We define the energy efficiency as follows:

$$\eta_{\mathrm{system}} = \frac{\dot{E}_{\mathrm{useful,HE}}}{\dot{E}_{\mathrm{brine}}} \tag{18}$$

The geothermal brine exergy input from the production field is calculated as follows:

$$\dot{E}x_{\text{brine}} = \dot{m}_{\text{w}}[(h_{\text{brine}} - h_0) - T_0(s_{\text{brine}} - s_0)] \quad (19)$$

The exergy destructions in the heat exchanger, pump, and the system itself are calculated using the following:

$$\dot{E}x_{\text{dest,HE}} = \dot{E}x_{\text{in}} - \dot{E}x_{\text{out}} = \dot{E}x_{\text{dest}}, \quad (20)$$

$$\dot{E}x_{\text{dest,pump}} = \dot{W}_{\text{pump}} - (\dot{E}x_{\text{out}} - \dot{E}x_{\text{in}}), \text{ and} \quad (21)$$

$$\dot{E}x_{\text{dest,system}} = \sum \dot{E}x_{\text{dest,HE}} + \sum \dot{E}x_{\text{dest,pump}} \quad (22)$$

We define the exergy efficiency as follows:

$$\psi_{\text{sys}} = \frac{\dot{E}x_{\text{useful,HE}}}{\dot{E}x_{\text{brine}}} = 1 - \frac{\dot{E}x_{\text{dest,sys}} + \dot{E}x_{\text{reinjected}} + \dot{E}x_{\text{natural discharged}}}{\dot{E}x_{\text{brine}}} \quad (23)$$

In this study, the reference environment was taken to be the state of environment at which the temperature and the atmospheric pressure are 13.1°C and 101.325 kPa, respectively, which were the values measured at the time when the GDHS data were obtained. For analysis purposes, the actual data were taken from the BGDHS on January 1, 2003, and the respective thermodynamic properties were obtained based upon these data. It is important to note that the number of the wells in operation in the IBGF may vary depending on the heating days and operating strategy.

Using Eq. 17, the total geothermal reinjection fluid mass flow rate is 111.02 kg/s at an average temperature of 66.1°C, and the production well total mass flow rate is 148.19 kg/s, and then the natural direct discharge of the system is calculated to be 37.17 kg/s on January 1, 2003. This clearly indicates that in the BGDHS, a significant amount of hot water is lost through leaks in the hot-water distribution network.

The exergy destructions in the system particularly occur in terms of the exergy of the fluid lost in the pumps, the heat-exchanger losses, the exergy of the thermal water (geothermal fluid) reinjected, and the natural direct discharge of the system, accounting for 3.06, 7.24, 22.66, and 24.1%, respectively, of the total exergy input to the BGDHS. Both the energy and the exergy efficiencies of the overall BGDHS are investigated for system performance analysis and improvement, and are determined to be 37.60 and 42.94%, respectively.

In the GDHSs, the temperature difference between the geothermal resource and the supply temperature of the district heating distribution network plays a key role in terms of exergy loss. In fact, the district heating supply temperature is determined after the optimization calculation. In this calculation, it should be taken into account that increasing the supply temperature will result in a reduction of investment cost for the distribution system and the electrical energy required for pumping stations, while it causes an increase of heat losses in the distribution network. Unless there is a specific reason, the district heating supply temperature should be higher to increase the exergy efficiency of the heat exchangers and, hence, the entire system. Besides this, in the design and operating condition of the primary heat exchangers, a temperature approach of about 3°C is desired. On the other hand, dropping the district heating supply temperature increases the amount of building heating equipment to be oversized. Oversizing does not mean only cost, but also more exergy production due to unnecessarily inflated pumping, pipe frictions, etc. In this regard, there is an optimum district flow rate and the minimum possible exergy loss (mainly due to pumping), of which determination is planned as a further future work to be conducted.

CONCLUSIONS

We have presented some historical background on DHC, along with cogeneration and GDHS applications, and discussed some technical, economical, environmental, and sustainability aspects of geothermal energy and performance evaluations tools in terms of energy and exergy analyses for such DHC systems. We also presented two case studies to highlight the importance of exergy use as a potential tool for system analysis, design, and improvement. The benefits have been demonstrated of using the principles of thermodynamics via exergy to evaluate energy systems and technologies as well as environmental impact. Thus, thermodynamic principles—particularly the concepts encompassing exergy—can be seen to have a significant role to play in evaluating energy and environmental technologies.

For societies to attain or try to attain sustainable development, effort should be devoted to developing DHC applications and technologies, which can provide an important solution to current environmental problems, particularly if renewable energy resources are used. Advanced renewable energy technologies can provide environmentally responsible alternatives to conventional energy systems, as well as more flexibility and decentralization.

REFERENCES

1. Spurr, M. District Energy/Cogeneration Systems in U.S. Climate Change Strategy. Climate Change Analysis Workshop, June 6–7, Springfield: Virginia, 1996.
2. IEA. Implementing Agreement on District Heating and Cooling Including the Integration of CHP. Strategy document, International Energy Agency: Paris, 2004.
3. Strickland, C.; Nyboer, J. Cogeneration Potential in Canada-Phase II. Report for Natural Resources Canada, 2002; 40, 6.

4. UNEP. Cogeneration, Energy Technology Fact Sheet, Division of Technology, Industry and Economics-Energy and Ozone Action Unit. United Nations Environment Programme, Nairobi, Kenya. Available at: www.uneptie.org/energy. Accessed on 14th November 2006.
5. Enwave. History of District Energy, 2006. Available at: www.enwave.com/enwave/view.asp?/solutions/heating/history. Accessed on 14th November 2006.
6. Dincer, I. Renewable energy and sustainable development: a crucial review. Renew. Sust. Energy Rev. **2000**, *4* (2), 157–175.
7. Dincer, I.; Rosen, M.A. Thermodynamic aspects of renewables and sustainable development. Renew. Sust. Energy Rev. **2005**, *9* (2), 169–189.
8. Dincer, I. The role of exergy in energy policy making. Energy Policy **2002**, *30* (2), 137–149.
9. Rosen, M.A.; Le, M.N.; Dincer, I. Thermodynamic assessment of an integrated system for cogeneration and district heating and cooling. Int. J. Exergy **2004**, *1* (1), 94–110.
10. Rosen, M.A.; Le, M.N.; Dincer, I. Efficiency analysis of a cogeneration and district energy system. Appl. Therm. Eng. **2005**, *25* (1), 147–159.
11. Edmonton Power. City of Edmonton District Energy Development Phase. Section 2: Engineering Report, 1991.
12. MacRae, K.M. *Realizing the Benefits of Community Integrated Energy Systems*; Canadian Energy Research Institute: Calgary, Alberta, 1992.
13. Bloomquist, R.; Nimmons, J. Geothermal district energy. In *Course on Heating with Geothermal Energy: Conventional and New Schemes* (Conveyor: Lienau, P.J.). WGC2000 Short Courses, June 8–10, 2000, Kazuno, Tohoku District, Japan, 2000; 85–136.
14. Hepbasli, A.; Canakci, C. Geothermal district heating applications in Turkey: a case study of Izmir–Balcova. Energy Convers. Manage. **2003**, *44*, 1285–1301.
15. Ozgener, L.; Hepbasli, A.; Dincer, I. Thermo-mechanical exergy analysis of Balcova geothermal district heating system in Izmir–Turkey. ASME J. Energy Res. Tech. **2004**, *126* (4), 293–301.

Drying Operations: Agricultural and Forestry Products

Guangnan Chen
Faculty of Engineering and Surveying, University of Southern Queensland, Toowoomba, Queensland, Australia

Abstract
This entry presents an overview of the methods for drying agricultural and forestry products. The need for the drying of agricultural and forestry products is presented, and the principles of drying operations are described. It is shown that a diversity of drying systems are currently used. The performance of these dryers varies significantly. Significant progress has been made in the research and development of drying technology to improve product quality, reduce cost, and improve environmental performance.

Dem—Dry

INTRODUCTION

Drying, the removal of water from products, is a necessary operation in many industries for the purpose of preserving product quality or adding value to the products. This entry focuses on the drying of agricultural and forestry products; in particular, grain drying and (solid) wood drying, which form the bulk of drying operations in these two industries.

THE NEED FOR DRYING OF AGRICULTURAL AND FORESTRY PRODUCTS

The main purpose of agricultural drying is often to reduce field loss and weather damage, and to maintain product quality during subsequent storage and delivery. This is in contrast to the drying of wood, whose main purpose is to have all the shrinkage and distortion take place before the wood is put into use. The reduction of insect and fungal attack is another reason for the need for wood drying.

In addition to the above difference, agricultural drying is also typically a seasonal activity, while wood drying is normally a year-long operation. This difference can have a significant impact on a number of aspects of drying operation and dryer design.

Overall, drying may be regarded as a risk management tool for the agricultural industry, and a value-adding tool for the forestry industry.

METHOD OF THERMAL DRYING

In most cases, drying involves the application of thermal energy. This is achieved by heating up the product and forcing hot air through it, therefore vaporizing and removing the moisture inside the product. In addition to the promotion of heat and mass transfer, the circulation of air also helps to carry the heat to and the moisture away from the product.

Significant energy is required in the drying process for several reasons:

- Raising the temperatures of air, the product, and water
- Vaporizing the water
- Compensating heat loss through radiation, convection, and operational losses (e.g., leaks)
- Compensating heat loss through the venting of heated humid air

Thermal drying, which involves water phase change, is a very energy intensive activity. For example, evaporating one cup (250 mL) of water would require approximately the same amount of energy as it would to heat a big pan of soup (3 L) from 25 to 70°C. Thus, the efficiency of energy use in drying processes is significant in the context of energy, economic, and environmental policy goals.

Because artificial drying normally offers the advantages of better control over product quality and higher productivity, this method has been widely used in the agricultural and forestry industries. Most artificial dryers also use the direct heat and vent method to drive the drying process.

HOW DRYING TAKE PLACE

During drying, evaporation may take place in two stages.[1,2] At first, there may be sufficient moisture within the product to replenish the moisture lost at the surface, until the critical point is reached and a dried surface forms. Evaporation is then principally dependent upon the rate of internal moisture diffusion. This is called the falling rate period or second period of drying, and is often a diffusion

Keywords: Grain drying; Wood drying; Drying performance; Drying process; Dryer design; Agricultural and forestry products; Product quality; Process energy efficiency; Drying technology.

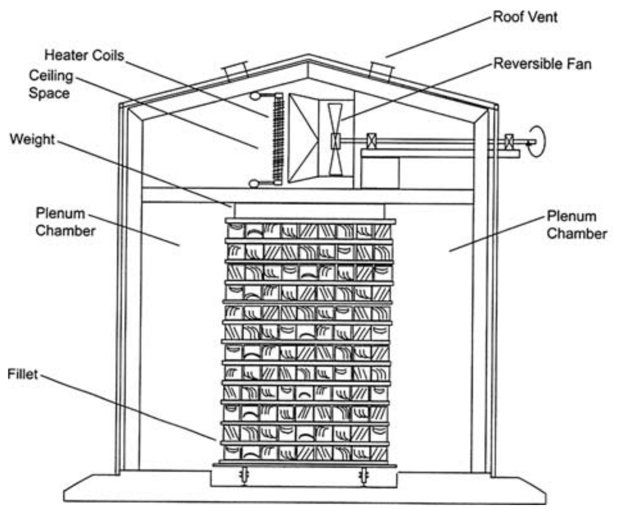


Fig. 1 Typical components and arrangement of a batch wood drying kiln.

process. Compared with the first period of constant drying at the rate of liquid water evaporation, diffusion is typically a slow process, and is mainly controlled by internal moisture transport of the product. Diffusion processes may be considerably accelerated with increased temperatures. External mass transfer plays a relatively small role at this stage.

Corresponding to the above process, initially, as the product surface is dried, it is restrained by the wet core so that it is subjected to a tensile stress. Later, as the core dries, it is in turn restrained by the drier surface, so that the stress profile inverts. At the end of drying, the product surface may be left with a residual compressive stress, whereas the core is subjected to a tensile stress. This is called case-hardening.[3] A drying schedule will therefore need to ensure that the stresses developed during any period of the drying process not exceed the strength of the material, so that stress damage of the product does not take place. At the end of drying, stress relief for the residual stresses may also be carried out.[4,5] This is particularly important when a product needs further processing or when a high-temperature fast drying schedule is employed. Cooling of agricultural products also minimizes the water condensation on the product surface.

TYPES OF DRYERS

A dryer generally consists of a chamber, a heating and air circulation system, and a control system.

Dryers may be classified in many ways, such as by modes of operation (e.g., batch or continuous dyers), by fuel sources, by drying temperature ranges, or by dryer throughputs. Further classifications are also possible, including heat transfer methods (e.g., direct or indirect heat transfer) and relative directions of the flows of product and air (e.g., cross-flow, counter-flow, and concurrent-flow).

In the agriculture and forestry sector, the most common type of dryer is still the fixed-bed batch (bin, shed, or compartment) dryer (Fig. 1). In these dryers, the product remains stationary while the drying environments are successively varied. This drying mode is particularly suitable for small to medium operators, with the advantage of low capital cost requirement. However, this method is also generally of lower capacity and more labor intensive. In comparison, a continuous dryer would typically use higher temperatures, have much larger capacity, and be more suitable for large operations. This method, however, requires large capital outlay. It is also sometimes more difficult to achieve accurate product specification with continuous driers, because of the potential impact of process air leakage and ambient conditions.

In the past decade, there has been an increasing interest in the use of various drying facilities, particularly low-cost, low-temperature dryers, as more and more farmers have begun to appreciate the importance of drying in the total harvesting system. The rapid growth of the forestry plantation industry also promotes the widespread installation of various timber-drying kilns.

Other specialist drying methods are also available. These include fluidized-bed drying for moist particulate products (such as grains, peas, and sliced vegetables) and drying by the application of energy from microwave or dielectric sources. However, many of these methods may only be cost effective for particular high-value products or for obtaining specific attributes for specific products. Freeze drying is reported to be able to achieve smaller product shrinkage, longer product storage life, and better retention of biological activity, so it is popular with the food industry.

DRYING SCHEDULES

A drying schedule may be described as a series of temperature, humidity, and air velocity settings used to dry the product to a specific moisture content, and to produce consistent, defect-free dry products in as short a time as possible with the least amount of energy use. Many agricultural and forestry products are required to be dried to a final moisture content of around 10%–12%.

There are two ways to define moisture content.[6] In general, the moisture content of forestry products is often expressed in dry basis (MC_{db}), which is the fraction of the mass of water in comparison with the mass of the oven-dry product. By contrast, the moisture content of agricultural products is normally expressed in wet basis (MC_{wb}), which is the amount of water in the product divided by the total product weight. These two definitions of MC_{db} and MC_{wb} can be converted to each other by the relationship

$MC_{db} = MC_{wb}/(1 - MC_{wb})$.

During drying, agricultural and forestry products usually undergo considerable changes, including shrinkage, cracking (both externally and internally), nutrient loss, and color changes. By imposing harsh drying conditions, a high-temperature regime may bring in the benefit of shortened drying schedules. However, such a process may also increase the risk of quality degradation, so a right balance needs to be achieved between these two competing factors.

The required drying time varies greatly for different agricultural and forestry products, ranging from a few hours to several days or even several weeks, depending on the product characteristics and the specified quality grades. Many agricultural products also have to be dried within a certain time frame to avoid significant quality deterioration.[7]

At present, most commercial dryers operate at comparatively moderate drying schedules to avoid the risk of quality loss. Low-value and more permeable materials may be dried more rapidly. In general, it may be categorically found that a high-temperature, high-humidity schedule may be suitable for fruit products, while a low-temperature, low-humidity schedule would be good for high-value seed products. In the middle, grain and timber are reasonably robust and may be suitable for a high-temperature, low-humidity schedule.

Currently, simple staged temperature controls are preferred for drying agricultural and forestry products, particularly for grain drying. To minimize degrades, some additional pre- or post-treatments such as initial air drying and post-drying cooling and conditioning may also be employed. Additional quality control may be attained by presorting the material prior to kiln drying so that the properties of batches are relatively homogenous. Grain inverter or airflow reversal may also be adopted.

Current drying schedules have been largely derived from trial-and-error experiments over a number of years. However, this method could be lengthy and expensive. Recently, a number of theoretical models have been developed to simulate and optimize the drying process and to reduce the number of laboratory and field experiments.[8,9]

Typical drying schedules for several agricultural and forestry products are as follows:

Permeable softwoods (such as radiate pine) for structural uses:

- Dry straight through from green to an average moisture content of 4% at a dry-bulb temperature of 120°C, a wet-bulb temperature of 70°C, and an air velocity of 5 m/s.
- Cool outside under cover for 90 min.
- Steam for 2 h.
- Cool with weight on and de-fillet within 24 h of steaming.

Permeable softwoods (such as radiate pine) for furniture uses (appearance grades):

- Dry at a dry-bulb temperature of 90°C and a wet-bulb temperature of 60°C, with a total duration of 2–3 days. Final steaming is also required, in order to remove the residual stress generated during the drying process. This is necessary, as the dried timber will be further reprocessed during furniture making.

Different from the softwoods, hardwoods are generally less permeable and more difficult to dry, so they are usually kiln dried by moisture content schedules.[10] This means that the dry- and wet-bulb temperatures are changed when the timber (lumber) reaches certain moisture contents. Hardwood is also often air dried first, before the kiln drying. Depending on the species and thickness of the lumber, the drying times may vary from one to a few weeks, with the final temperature being gradually raised from the ambient temperature to between 45 and 65°C. The air velocity is typically maintained at 1–1.5 m/s.

For grain drying (milling grade), it is generally recommended that the maximum drying temperature be limited to 70°C to minimize the heat damage, particularly at initial period of high moisture content. Feedstock grades can be dried at much higher temperatures. Seed drying is usually limited to 30°C–40°C. The common airflow rate for grain dryers is in the range of 200–1000 L/s/t of grain.

DRYER PERFORMANCE AND ENERGY EFFICIENCY

A dryer may be regarded as an energy system. Various energy sources may be used in the drying process, including electricity and various primary fuels such as coal, diesel, and gas. These energy sources all have different heating values, costs, and environmental impacts. Although electricity is a convenient and "clean" energy source, it is a high-grade energy because the typical efficiency of thermal generation of electricity is only 35%–50%. Overall, electricity is generally more expensive, particularly after taking account of the associated supply and transmission charges.

Typically, the energy cost for a small or medium drying operation may range from a few thousand dollars to over twenty thousand dollars; depending on and significantly influenced by the quantity and initial moisture content of the product, and operation practice such as drying schedules and controls. Assuming a 5% moisture removal, the total amount of water being removed from one ton of product is about 50 kg. Currently, in a commercial dryer, the energy required to evaporate 1 kg of moisture from a product ranges from 3.5 to 7.0 MJ.[11,12]

In many cases, it has been found that there is little correlation between the dryer energy performance and product process requirement. Lower process requirements do not necessarily lead to higher energy efficiency. This indicates that there is a significant potential to improve the dryer energy performance.

Although the technology is currently available, it is noted that there are significant barriers for the uptake of energy-efficient technology in the drying industry. This is because present production methods have historically been based on considerations of process throughput, reliability, and capital cost. Energy costs, although comprising a significant part of total operating costs in the drying of agricultural and forestry products, typically represent only 2%–5% of product value, and are therefore often of low priority. This is further reinforced by the factor that drying may be a secondary activity for many farmers and operators. This is particularly the case for agricultural dryers, as agricultural drying is typically a highly seasonal activity. Most agricultural dryers are only utilized for one to three months.

DRYER DESIGN AND SELECTION

Established procedures are now available for the design of agricultural and forestry product dryers. In spite of this, the actual performance of different industries and different dryers still varies considerably. Poor drying systems can lead to significant penalties in terms of lost production and lost income, including increased energy cost and degraded or non-uniform product. In comparison with the "seasonal" agricultural drying industry, the timber industry is typically a year-long operation, and hence more expensive technology and personnel training may be justified. For example, in the timber drying industry, automatic kiln monitoring, management, and control systems have now been routinely implemented to improve the dryer performance. This is relatively rare for the drying of agricultural products. Because of the short period used, most agricultural crop dryers are also not insulated.

To obtain the maximum performance and the desired product quality, it is important that a suitable drying system be selected, with correct system sizing, matching of subsystems, and operating procedures. These factors are often interlinked, so an integrated and holistic approach is required. For example, when employing a high-temperature regime, more water vapor is produced, and higher airflow rates will be required to carry away this vapor. When contemplating increasing the dryer capacity or adopting a new dryer, it is also important to consider the impact on all the other components of crop harvesting and storage systems. A decision should not be made on the basis of the effects of that particular facility alone.

To save energy, some large sites may have several dryers so that a cascade arrangement for exhaust energy recovery is justified. For a large production, it may be possible to carry out electricity tendering or form an electricity user "club" to reduce the electricity tariff. Optimization of fan sizing and operation is also important, as fan laws stipulate that a 50% fan speed reduction can result in a reduction of fan power to only one-eighth of the original power requirement. Less fan power may be needed during the later stages of the diffusion drying process.

At present, a number of computer models have been developed and used in the dryer design. The main advantage of this method is to achieve a precision sizing of the equipment and to produce a predictable design to reduce the customer's business risk. For agricultural drying, climate-based models[13] have also been developed to ensure optimal design and integration between various agricultural machinery, crop performance, and perceived weather risk. Together with local historical weather data and future climate forecasts, these models have been used to assist in the decisions of both long-term investment in drying facilities and short-term tactical operation decisions (e.g., by adjusting the crop planting and harvesting schedule, by crop choice and crop diversification, or by early negotiation with harvest and drying contractors).

RECENT RESEARCH AND DEVELOPMENT

Significant research has been carried out in the area of agricultural and forestry product drying. This has included research on dryer design, the impact of drying on product quality, improvements of drying energy efficiency, new methods of drying, and applications of new technologies.

Dryer Design and Operation

Because most of the current agricultural dryers are of small to medium throughputs and are operated by rural family businesses, the main constraints for agricultural drying are often the capital expenditure and the technical competence and skills of farmers and local dryer manufacturers. It has been found that the common reasons for poor dryer performance are poor dryer design, inappropriate equipment selection and installation, and bad control and operating practice. Practical information on the best operating practice also tends to be fragmented and not readily available in a useable form.

One of the current research priorities is therefore to demonstrate and establish appropriate technologies to overcome the above barriers and to improve the integration of heat and mass transfer processes and the matching of subsystems.[14]

For commercial operators, uneven drying is also a significant problem. This may be difficult to eliminate, because drying environment is inherently dynamic and varies with locations inside the dryer. Furthermore, dryers may also be required to handle variable resources, including feed materials of different species, non-uniform initial moisture contents, and different sizes. Computational fluid dynamics (CFD) has now been widely used to improve the dryer design and to minimize air recirculation loss. The latest research is also focusing on the development of new sensor techniques and enhanced machine vision tools for collecting quality control information and developing expert systems for rapid problem diagnosis.

Considerable effort is also being made to investigate the effect of drying conditions on the shrinkage, stress development, and quality of products.

New Methods of Drying and New Technologies

A number of innovative methods and drying technologies are being developed. Among them, drying with a modified atmosphere has shown significant promise. Drying and storing fruit and foods in a controlled atmosphere (CA) can lead to improved product quality, because displacing oxygen with other gases such as N_2 and CO_2 retards the oxidation process. Considerable commercial success has been achieved in the area of CA fruit storage and transport. Similar opportunities have also been identified in the area of fruit drying, particularly in terms of eliminating the use of chemical preservatives or other additives. Recent experiments have shown that the use of innovative CA, oxygen-free drying for apples can significantly improve the product attributes, in particular reducing brown staining and avoiding the requirement of using sulfur chemical pretreatment.[15] This can lead to more healthy products, with the additional benefits of improved taste (no acid) and better texture.

Since drying is an energy-intensive operation, much attention is also given to the development of an energy-efficient drying process. Heat-pump drying is one such technology, because the heat normally vented to the atmosphere is recovered. A heat-pump dryer (HPD) is essentially an industrial adaptation of a normal air conditioning system. Energy (electricity) inputs to the dryer include those to the compressor and the fans. For each unit of electrical energy used by the heat pump, generally three to four units of energy are available for drying the product. In a heat-pump dehumidifier, most of the moisture is also removed from the kiln as liquid rather than moist warm air.

Due to the limits of currently applied working fluids (refrigerants), the HPD normally has to operate at low to medium temperatures, so the drying rates are also slower. This is suitable for a number of heat-sensitive products, but also makes it difficult to compete with alternative mainstream technologies, where the emphasis is often put on the fast drying rate and quick return of plant capital costs.[16] Solar-assisted HPDs are also being developed[17] to accelerate the drying process.

CONCLUSION

Drying is a significant operation in the agricultural and forestry industries. Considerable progress has been made in the research and development of drying technology to improve product quality, reduce cost, and improve environmental performance. It has been shown that a diversity of drying systems are used. The efficiency and performance of different driers vary significantly. Energy consumption is strongly influenced by the dryer design, the particular operation practice, and the individual skills of the operator.

A number of innovative methods and drying technologies are being developed. Among them, drying with a modified atmosphere and improvements in sensor and control technology have shown significant promise.

REFERENCES

1. Mujumdar, A.S., Ed. *Handbook of Industrial Drying*; Marcel Dekker: New York, 1989.
2. Keey, R.B. *Drying of Loose and Particulate Materials*; Hemisphere Publishing Corporation: New York, 1992.
3. Keey, R.B.; Langrish, T.A.G.; Walker, J.C.F. *Kiln-Drying of Lumber*; Springer: Berlin, 2000.
4. Pratt, G.H. *Timber Drying Manual*; Building Research Establishment: London, 1974.
5. Haslett, A.N.; Simpson, I. Final steaming critical to HT drying of appearance-grade timber. NZ Forest Industries **1991**, (9), 40–41.
6. Broker, D.B.; Bakker-Arkema, F.W.; Hall, C.W. *Drying and Storage of Grains and Oilseeds*; AVI Publishing Co.: New York, 1992.
7. Raghavan, G.S.V. Drying of agricultural products. In *Handbook of Industrial Drying*; Mujumdar, A.S., Ed.; Marcel Dekker: New York, 1987; 555–570.
8. Stanish, M.A.; Schajer, G.S.; Kayihan, F. A mathematical model of drying for hygroscopic porous media. AIChE J. **1986**, *32* (8), 1301–1311.
9. Wu, Q.; Milota, M.R. Rheological behaviour of Douglas-fir at elevated temperature. Wood Fibre **1995**, *27* (3), 285–295.
10. Simpson, W.T., Ed. *Dry Kiln Operator's Manual*, *Agriculture Handbook No. 188*; U.S. Department of Agriculture: Washington, DC, 1991.
11. Chen, G.; Anderson, J.A.; Bannister, P.; Carrington, C.G. Monitoring and performance of a commercial grain dryer. J. Agric. Eng. Res. **2001**, *81* (1), 73–83.
12. Hill, M., Ed. *Drying and Storage of Grains and Herbage Seeds*; Foundation of Arable Research: Lincoln, New Zealand, 1999.
13. Abawi, G.Y. Asimulation model of wheat harvesting and drying in northern Australia. J. Agric. Eng. Res. **1993**, *54*, 141–158.

Dem–Dry

14. Carrington, C.G.; Sun, Z.F.; Sun, Q.; Bannister, P.; Chen, G. Optimising efficiency and productivity of a dehumidifier batch dryer. Part 1—Capacity and Airflow. Int. J. Energy Res. **2000**, *24*, 187–204.
15. Perera, C.O. Modified atmosphere drying. Chapter 6. In *Drying of Products of Biological Origin*; Mujumdar, A.S., Ed.; Oxford and IBH Publishers: New Delhi, India, 2004; 153–163.
16. Bannister, P.; Carrington, C.G.; Chen, G. Heat pump dehumidifier drying technology—status, potential and prospects. Proceedings of 7th IEA (International Energy Agency) Heat Pump Conference, Vol. 1, 219–230, China Architecture and Building Press: Beijing, 2002.
17. Chen, P.Y.S.; Helmer, W.A.; Rosen, H.N.; Barton, D.J. Experimental solar-dehumidification kiln for drying lumber. Forest Prod. J. **1982**, *32* (9), 35–41.

Drying Operations: Industrial

Christopher G. J. Baker
Chemical Engineering Department, Kuwait University, Safat, Kuwait

Abstract
Dryers are widely used on an industrial scale and are major consumers of energy. This article first briefly reviews some of the more common types of drying equipment. It then goes on to discuss the energy consumption of dryers, with particular emphasis on an ideal adiabatic dryer model against which the performance of drying equipment in the field can be benchmarked. It will then describe the use of energy audits to quantify potential energy savings. Finally, it will discuss low-cost and capital-intensive schemes for reducing the energy consumption of dryers.

NOMENCLATURE

C_{pg} Specific heat of dry air (kJ/kgK)
E_s Specific energy consumption (GJ/t)
$E_{s,a}$ Specific energy consumption of adiabatic dryer (GJ/t)
E_s^* Measured specific energy consumption (GJ/t)
F Feedrate (dry solids basis) (kg/s)
Q_{ev} Heat required to provide the latent heat of evaporation (kW)
Q_{htr} Heat supplied to dryer by heater (kW)
T Temperature (°C)
W_{ev} Evaporation rate (kg/s)
X Moisture content (mass water per unit mass of dry solids) (kg/kg)
Y Humidity (mass of water vapor per unit mass of dry air) (kg/kg)

Greek letters
η Thermal efficiency of dryer (defined by Eq. 1) (%)
λ_{ref} Latent heat of evaporation of water at 0°C (GJ/t)
ξ $(T_o - T_a)/(Y_o - Y_a)$ (°C)

Subscripts
a Ambient
i Inlet
o Outlet

INTRODUCTION

Drying can be defined as that unit operation which converts a liquid, solid, or semi-solid feed material into a solid product that has a significantly lower moisture content. Although there are some notable exceptions, drying is normally achieved through the application of thermal energy, which is used in part to supply the latent heat of the evaporation of water. In certain industrial processes, drying may also involve the removal of organic solvents, either alone or in combination with water.

Drying forms an integral part of many industrial manufacturing processes. Examples can be found in the following sectors: chemicals, petrochemicals, polymers, food, agriculture, pharmaceuticals, ceramics, minerals, paper and board, textiles, etc. Dryers come in a wide variety of configurations with throughputs ranging from 50 kg/h or less to tens of tonnes per hour. As would be anticipated, batch dryers are employed at relatively low throughputs and continuous dryers at higher throughputs. The exact demarcation between these two categories of dryer is based on a number of technical and economic factors.

The principal types of batch and continuous dryers are classified into different categories as shown in Figs. 1 and 2, respectively. In these figures, the term "layer dryer" is used to describe those devices in which a surface within the dryer is employed to support and/or heat the coherent mass of drying solids. Conversely, in dispersion dryers, the solids are freely suspended in the hot air flow. The feedstock may be heated directly by convection with hot air, as in dispersion dryers, or indirectly by conduction through a heat-exchange surface, as in contact dryers. In the latter case, operation is possible in a vacuum as well as at atmospheric pressure. Special dryers include those in which energy is wholly or partially supplied by means of dielectric (microwave and radiofrequency) heating or water is removed by sublimation, as in freeze dryers. Both types are relatively expensive in terms of capital and operating costs. Therefore, the use of dielectric dryers is largely restricted to the removal of relatively small traces of moisture, which is expensive by conventional means. Freeze drying can be employed when low-temperature processing enhances a product's quality and value, as is the case with freeze-dried coffee.

A selection of typical and widely used industrial dryers is described in Table 1 and illustrated schematically in

Keywords: Dryers; Dryer selection; Dryer operation; Energy consumption; Energy auditing; Energy conservation.

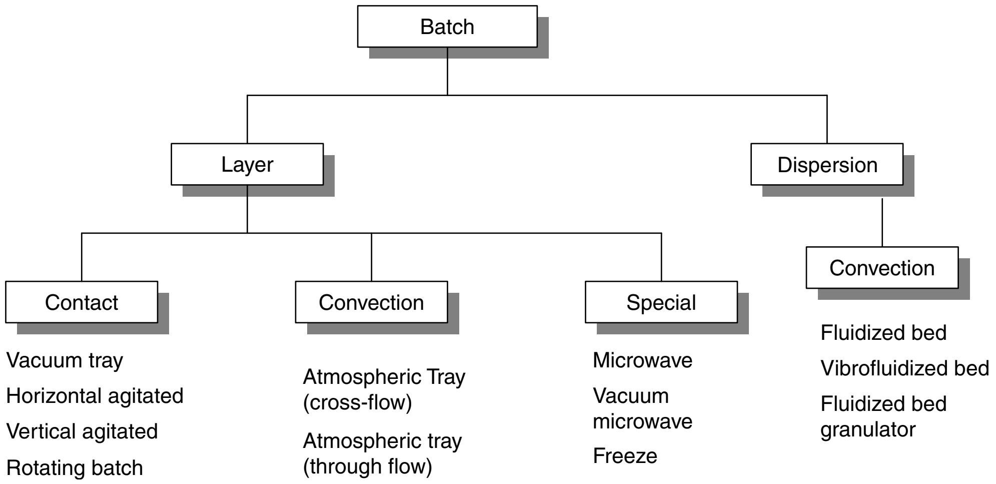


Fig. 1 Simplified classification of batch dryers.

Figs. 3–7. Because of the complexity of drying operations, many factors have to be considered and weighed when selecting an appropriate dryer for a given application.[1] Often there is no one "right" answer as several options may be both technically and economically viable. The optimal choice can be defined as that dryer which satisfies all process requirements at minimum cost. Process requirements may include the specification of designated quality parameters in addition to exit moisture content. Minimum cost is often taken to mean minimum capital cost, but this ignores the expenditures on fuel, which dominate the cost of the drying operation over its economic life.

Drying processes involve simultaneous heat and mass transfer, the underlying theory of which is relatively complex and falls beyond the scope of this article. The interested reader is therefore referred to one of a number of specialized handbooks on drying[2,3] or to the appropriate chapter in a more general Chemical Engineering textbook.[4,5] This article concentrates primarily on those aspects of dryer operation that impact on their energy use.

ENERGY CONSUMPTION OF DRYERS

Drying processes consume very large quantities of energy. There are several reasons for this. The first, as noted above, is their widespread use throughout industry. Secondly, by their very nature, dryers are highly energy intensive. As a rule of thumb, a typical convective dryer

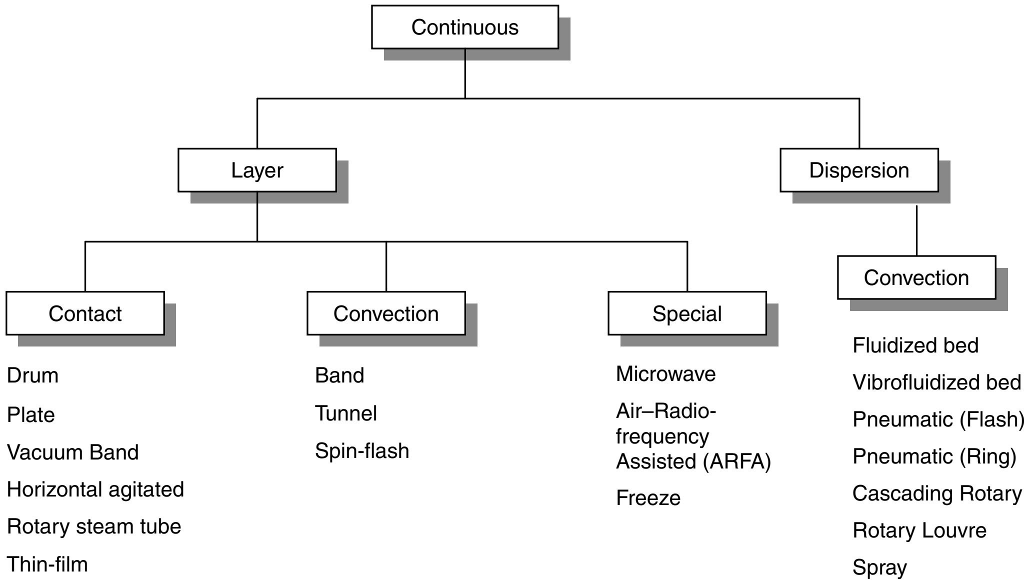


Fig. 2 Simplified classification of continuous dryers.

Table 1 Examples of different dryer types

Type	Description	Feedstocks and products
Horizontal agitated dryer (Several configurations, batch or continuous operation)	Horizontal cylindrical heated shell with axial paddles. Atmospheric or vacuum operation	Solids, pastes, or solutions. Used to dry, e.g., products ranging from foodstuffs to pigments. Low throughputs
Drum dryer (Several configurations, continuous operation)	Slowly rotating hollow drum heated internally by steam. Continuous operation. Most models operate at atmospheric pressure; vacuum model available	Solutions and pastes. Products range from instant potatoes to salts. Low throughputs
Spray dryer (Several configurations, continuous operation)	Used for drying liquid feeds, which are atomized into small droplets, and contacted with hot gas in a large drying chamber	A large number of products including liquid foods (e.g., coffee, milk), chemicals, plastics, etc. Capable of processing very high throughputs
Fluidized bed dryers (Several configurations, batch or continuous operation)	Floating bed of solids supported by stream of hot drying air	Widely used to dry small, free-flowing solids—foods, minerals, chemicals, plastics, etc. Can be used to dry/agglomerate particles. High throughputs
Rotary dryers (Several configurations, continuous operation)	Large, slowly rotating sloping drum. In one version, solids are picked up by flights on the periphery and showered through a hot air stream	Very high throughputs of solid products ranging from minerals to chemicals, fertilizers and food by-products

can be expected to consume at least 1 MW of thermal power per t/h of evaporation. Finally, as dryers are frequently operated very inefficiently, this figure may in practice be considerably higher than it needs to be. As a result, many companies view dryers as popular targets in their energy conservation programs. The recent escalation in the price of oil and natural gas has naturally provided an added incentive. Additionally, the Kyoto Protocol has focused the need for governments of signatory countries to take drastic actions to curb

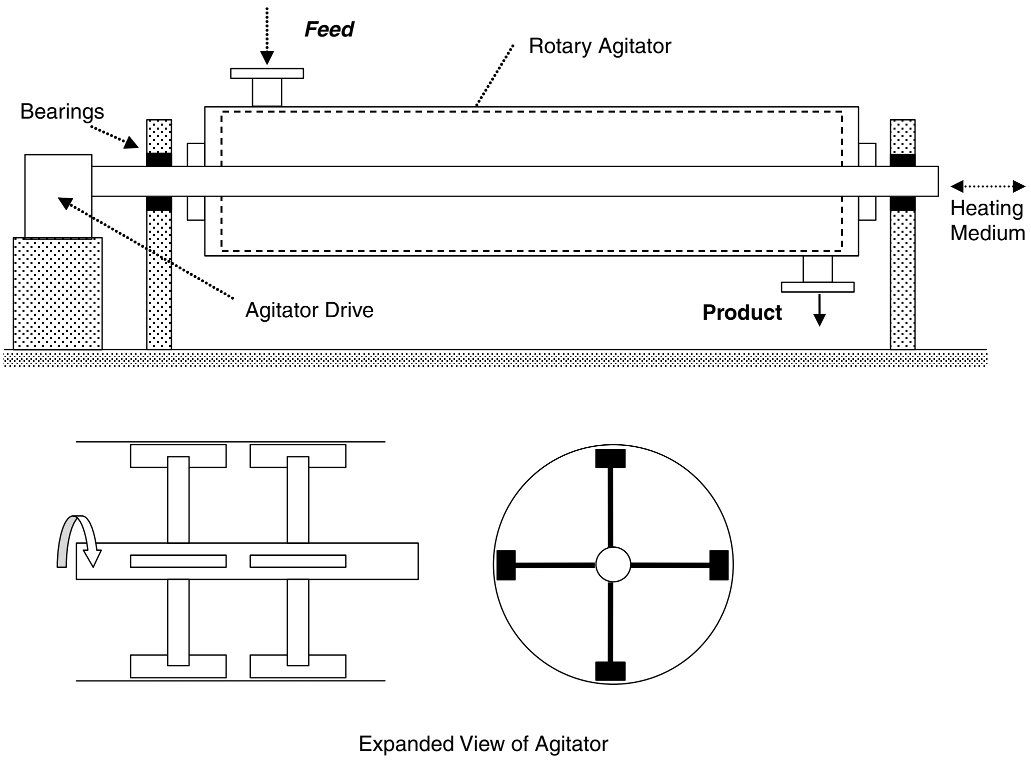


Fig. 3 Horizontal agitated dryer.

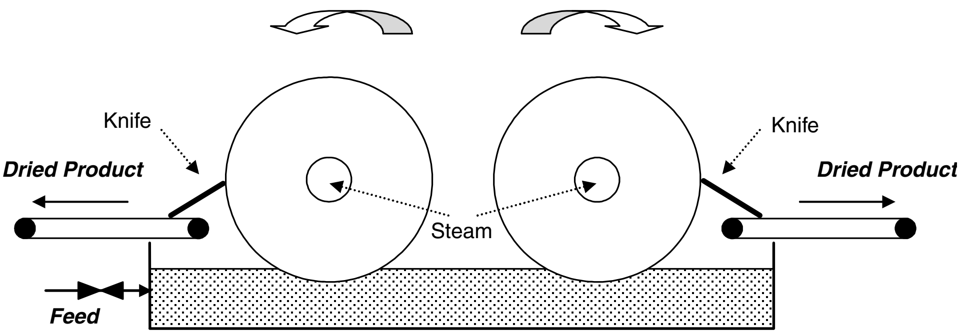


Fig. 4 Drum dryer (Double drum, dip feed).

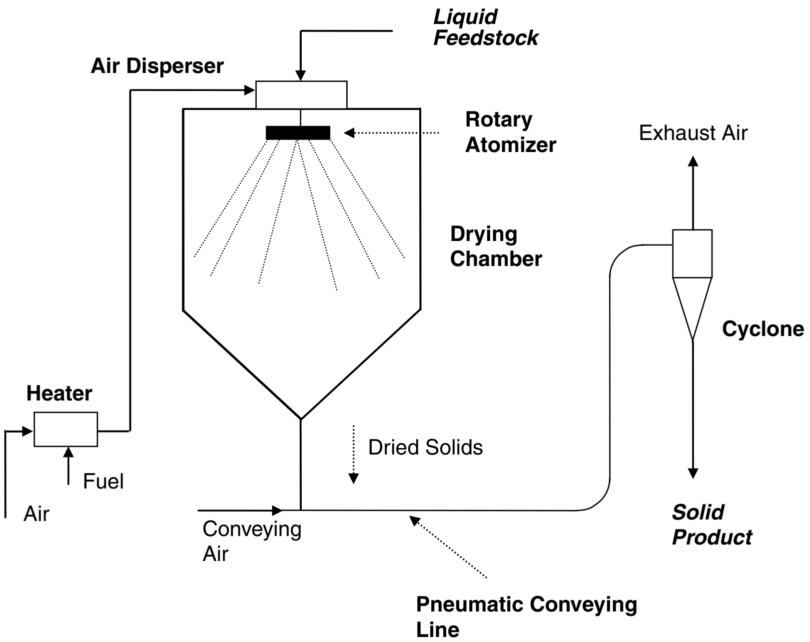


Fig. 5 Cocurrent spray dryer with rotary atomizer and pneumatic conveying of dried powder.

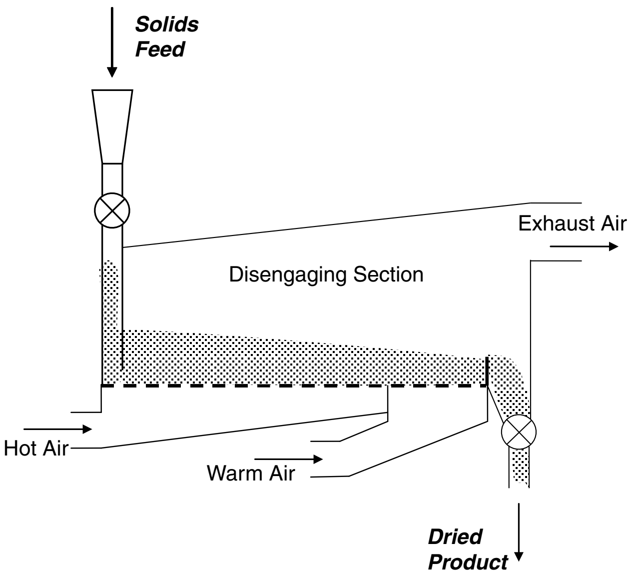


Fig. 6 Plug flow fluidized bed dryer.

greenhouse gas emissions. In order to achieve this, a body of enabling legislation has been introduced, particularly within the European Union. This legislation and its likely impact on both dryer manufacturers and operators is reviewed elsewhere.[6]

The most recent analysis of dryer energy consumption within the United Kingdom was published by Gilmour et al.[7] and includes a number of interesting statistics. For example, in 1994, estimates of the energy consumed in drying ranged from 348.6 to 379.5 PJ, depending on the method of calculation employed; the corresponding figure for the total industrial energy consumption in that year was 1969 PJ. Moreover, their analysis suggests that the proportion of energy consumed in drying progressively increased from around 11.6% in 1982 to 17.7%–19.3% in 1994. Another interesting statistic cited by Gilmour et al. is that the cost of fuel consumed by a typical convective dryer over its lifetime will likely exceed five times its initial capital cost. Given the recent worldwide

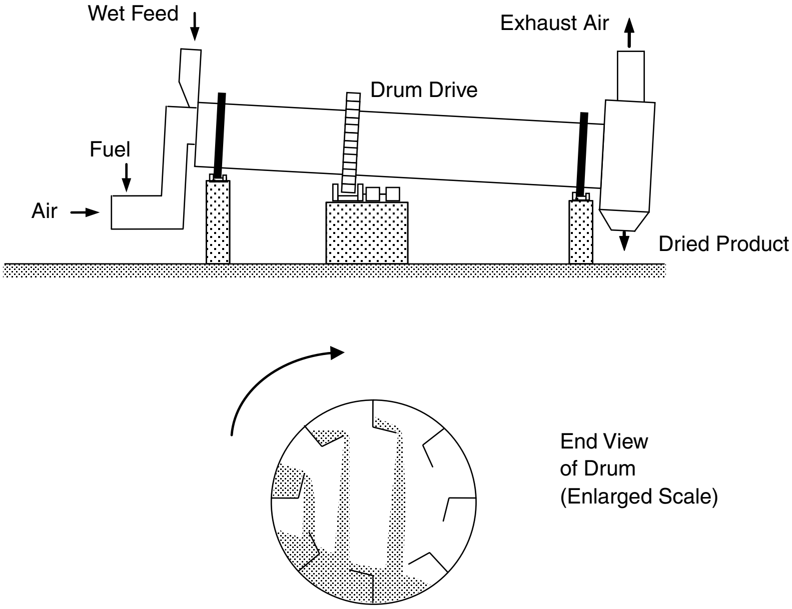


Fig. 7 Cascading rotary dryer with cocurrent solids-air flow.

escalation in oil and natural gas prices, this figure is probably a significant underestimate in today's terms.

The thermal efficiency of a dryer can be expressed in several ways. A typical measure is:

$$\eta = 100 \frac{Q_{ev}}{Q_{htr}} \tag{1}$$

where Q_{htr} is the total rate at which thermal energy is supplied to the dryer. Of this, Q_{ev} is required to provide the latent heat of evaporation. Alternatively, the specific energy consumption E_s of the dryer is defined as the thermal energy required to evaporate unit mass of water:

$$E_s = 0.001 \frac{Q_{htr}}{W_{ev}} \tag{2}$$

where W_{ev} is the evaporation rate. One might expect E_s to be approximately equal to the latent heat of evaporation of water, namely 2.5 GJ/t of water evaporated. In practice however, much higher values are observed (see below).

Gilmour et al.[7] cited previously published values of η for different industry sectors in the United Kingdom. As can be seen in Table 2, these differed quite widely, presumably reflecting the mix of dryer types employed and products dried, amongst other factors. Assuming that these figures are typical of those for other industrialized countries, the variation in the drying efficiencies cited suggests that there should be considerable scope for reducing dryer energy consumption. However, as will be discussed below, this statement has to be viewed with caution as many theoretical improvements may not be achievable in practice.

Baker and McKenzie[8] undertook a survey of the energy consumption of 32 industrial spray dryers in the ceramics, chemicals, and food industries. The survey, which was commissioned by the U.K. Government's Energy Efficiency Best Practice Programme, included dryers evaporating a total of 67.8 t/h of water. The thermal energy input to these dryers was 92.6 MW. The results of this survey, which included dryers having evaporation rates ranging from 0.1 to 12 t/h, revealed values of E_s varying from around 3–20 GJ/t of water evaporated. The average for all dryers included in the survey was 4.87 GJ/t.

The data obtained in the above survey were interpreted with the aid of a model that enabled the performance of a particular dryer to be compared with that of its ideal

Table 2 Dryer performance in selected industrial sectors

Industrial sector	Average drying plant efficiency, %	Total drying energy, PJ
Paper and board	50.0	91.8
Ceramics and building materials	69.3	79.0
Food and agriculture	47.1	123.0
Plaster and plasterboard	60.0	2.5
Textiles	57.3	38.7
Timber	55.0	9.6
Chemicals	58.0	3.3
Pharmaceuticals	70.0	0.02
Laundry	53.0	0.7

Source: From Gilmour et al. (see Ref. 7).

adiabatic counterpart. Baker and McKenzie showed that the specific energy consumption $E_{s,a}$ of such a dryer is not fixed in the absolute sense, but rather that it depended on the temperature and humidity of the outlet air:

$$
\begin{aligned}
E_{s,a} &= 0.001\left[C_{pg}\left(\frac{T_o - T_a}{Y_o - Y_a}\right) + \lambda_{ref}\right] \\
&= C_{pg}\xi + \lambda_{ref}
\end{aligned}
\tag{3}
$$

In this equation, C_{pg} is the specific heat of dry air, and λ_{ref} is the latent heat of water at 0°C. T_a and Y_a denote the temperature and humidity of the ambient air, and T_o and Y_o denote the corresponding values for the outlet air leaving the dryer. The latter are set so as to achieve the product's desired moisture content.

For an adiabatic dryer, it follows from Eq. 3 that a plot of $E_{s,a}$ against $\xi = (T_o - T_a)/(Y_o - Y_a)$ should be linear with a slope of 0.001 C_{pg} and an intercept of $0.001\lambda_{ref}$. This plot can be used as a baseline against which the performance of non-adiabatic dryers can be judged. Fig. 8 shows the data obtained in Baker and McKenzie's survey[8] plotted as E_s against $(T_o - T_a)/(Y_o - Y_a)$, in which T_a was taken as 25°C and Y_a as 0.005 kg/kg. As would be expected, all of the points scattered on or above Eq. 3, which depicts the performance of an ideal adiabatic dryer.

Eq. 3 can be used to interpret the performance of dryers in the field. As illustrated schematically in Fig. 9, a dryer exhibiting a given specific energy consumption may either be wasting considerable quantities of energy or operating close to peak efficiency. Here, the actual energy consumption of Dryer 1, E_s^*, is significantly larger than the value $(E_{s,a})_1$ required for an adiabatic dryer for which $\xi = \xi_1$. The efficiency of this dryer is relatively low. In contrast, for Dryer 2 ($\xi = \xi_2$), $E_s^* - (E_{s,a})_2 \ll E_s^* - (E_{s,a})_1$. This indicates that Dryer 2 is much more efficient than Dryer 1. Baker and McKenzie[8] found that the efficiency of the 32 spray dryers included in their survey varied widely. On average, though, almost 30% of the energy supplied to the dryers was wasted.

Specific energy consumption therefore provides a useful guide as to how efficiently (or not) a particular dryer is operating. The value of E_s must be measured by means of an energy audit on the dryer. Guidance on undertaking such audits is provided in the following section.

DRYER ENERGY AUDITS

Any attempt at reducing the fuel consumption of an industrial dryer should begin with assessments of how much energy is currently being consumed and whether this is being expended usefully or is being wasted. This establishes the benchmark against which future energy savings can be judged. In order to make the above assessments, it is first necessary to carry out an energy audit. This consists of a series of measurements designed to establish pertinent energy and mass flows into and out of the dryer. As each dryer is different, it follows that each audit should be customized to fit it.

There are essentially two types of audit, namely a basic audit and a detailed audit. In both cases, the dryer should be properly instrumented. Fig. 10 illustrates an ideal arrangement of measuring points. The primary purpose of a basic audit is to establish the specific energy consumption of the dryer and to assess the potential energy savings that are possible. If appropriate, a detailed audit can subsequently be undertaken to determine where the energy losses are

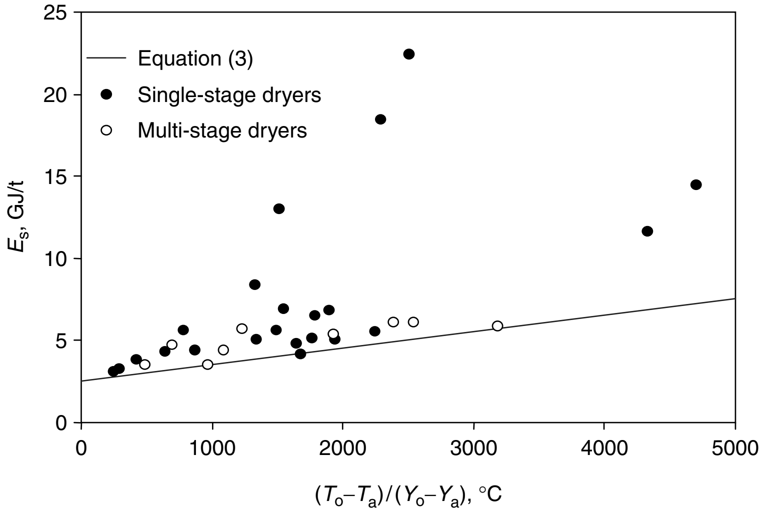


Fig. 8 Plot of specific energy consumption against $(T_o - T_a)/(Y_o - Y_a)$ for spray dryers (after Baker and Mckenzie[8]).

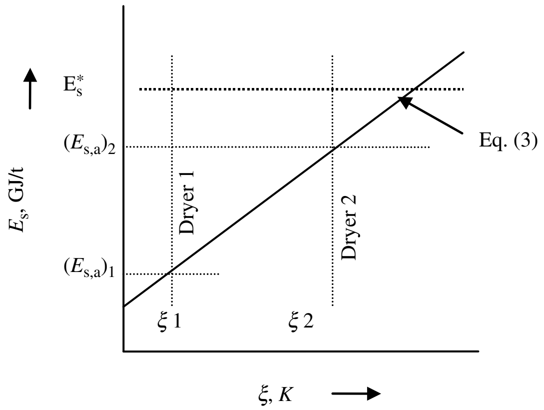


Fig. 9 Schematic representation of dryer efficiency.

occurring and to identify the corrective actions necessary to improve the performance of the dryer.

Process measurements on operating dryers are difficult to perform and the following precautions should be taken to minimize errors. All measurements should be made in duplicate or triplicate in order to ensure their accuracy and to confirm that the dryer is operating at steady-state. Triangulation, in which the value of a particular parameter is arrived at by more than one approach, is always advisable as it provides added confidence in the data. It is also important to ensure that all the measuring instruments employed are properly calibrated and used in accordance with recognized (e.g., ISO) standards to ensure accurate results.

Basic Audits

Table 3 lists the measurements that are required for a basic audit, the purpose of which is to determine E_s and ξ so that the performance of the dryer relative to its adiabatic counterpart can be assessed. The specific energy consumption can be calculated from the following equation:

$$E_s = \frac{0.001\ Q_{htr}}{F(X_i - X_o)} \tag{4}$$

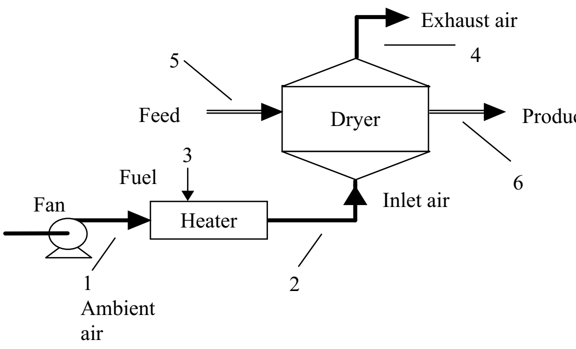


Fig. 10 Location of measurement points (after Baker[6]).

Table 3 Measurements required for a basic audit

Number	Measurement point	Measurements made
1	Ambient air	Air flowrate, temperature, and humidity
2	Inlet air	Air temperature
3	Fuel	Fuel flowrate[a]
4	Exhaust air	Air temperature and humidity[b]
5	Feed	Solids moisture content
6	Product	Solids flowrate, moisture content

[a]If steam is used as the heating medium, its temperature and pressure should be measured as well. Where possible, the condensate temperature and flow should also be recorded.
[b]It may be more appropriate to calculate this—see text.

in which F is the dry-basis solid feed rate, X_i and X_o are the moisture contents of the feed material and product, respectively, and Q_{htr} is as defined above. The measurement of F, X_i, and X_o is normally straightforward. Assuming that the dryer is operating at steady-state and that solid losses in, for example, the exhaust air, are minimal, F can be calculated from the production rate, which will undoubtedly be recorded on an ongoing basis for commercial reasons. Two methods are commonly used to determine the heat load Q_{htr}:

1. Direct measurement—This naturally requires that a fuel or steam meter be fitted.
2. Indirect measurement—This is useful in cases where a fuel meter is not fitted. One commonly used method is to carry out an energy balance over the heater.

The use of Method 2 presupposes that the air flowrate is known. This is normally measured using, for example, a Pitot tube. If this is not possible, the original dryer design specifications may, in some circumstances, give a reasonable estimate. However, there is clearly some uncertainty in the latter approach as the operating conditions may have changed considerably since the dryer was commissioned.

Methods 1 and 2 can of course both be used in the same audit. This is a good example of triangulation and increases confidence in the measured values of the energy supplied by the heater and the mass flowrate of the drying air.

In order to determine ξ, it is necessary to measure T_a, T_o, Y_a, and Y_o. The first three of these variables are straightforward. However, the measurement of Y_o is more problematical as the exhaust air is often dusty and its temperature and humidity are relatively high. Under these circumstances, the most accurate technique to determine

Table 4 Results of hypothetical basic audit on dryers having an evaporative load of 5 t/h

Dryer	ξ, °C	$E_{s,a}$, GJ/t	E_s^*, GJ/t	$E_s^* - E_{s,a}$, GJ/t	Potential saving, MW	Potential saving, %
A	6545	9.04	11.90	2.86	4.0	24.0
B	3333	5.83	9.47	3.64	5.1	38.6
C	2870	5.37	6.20	0.83	1.2	13.4

Y_o is often to calculate it from a mass balance over the dryer.

Table 4 shows some typical results that might be obtained through basic audits on three hypothetical dryers, each evaporating 5 t/h of moisture. In this table, the potential energy saving in MW is $W_{ev}(E_s^* - E_{s,a})$, where W_{ev} is the evaporation rate in kg/s and E_s^* and $E_{s,a}$ are in GJ/t. The numbers cited represent the maximum possible energy savings that can be achieved when the dryer is operating at the same exhaust air temperature and humidity.

Detailed Audits

Detailed audits are much wider in their scope than basic audits and may involve considerably more measurements. Their principal purpose is to identify the causes of inefficiencies in the dryer. This is accomplished by constructing detailed mass and energy balances around both the dryer and its air heater, as shown in Fig. 11. From these balances, it should be possible, for example, to determine the following:

- Magnitude of air leaks (if any) into or out of the dryer
- Magnitude of internal steam leaks (if any) within the heater
- Magnitude of the heat losses from the heater and dryer

Although interpretation of the data obtained in detailed audits requires a specialist's knowledge, they can, as discussed below, provide useful guidance on possible measures that can be employed to reduce a dryer's energy consumption.

PRACTICAL MEASURES FOR REDUCING DRYER ENERGY CONSUMPTION

Data obtained from detailed audits can be used to evaluate the effectiveness of various options for reducing the energy consumption of the dryer. These can be divided into (i) schemes involving little or no capital expenditure and (ii) schemes involving significant capital expenditure. A detailed quantitative analysis of the results of a typical dryer audit has been described by Baker.[6]

It should be stressed that any alterations to the dryer itself or to its operating conditions may involve an element of risk and should not be undertaken without a proper evaluation of all the factors involved. Where appropriate, expert guidance should be sought.

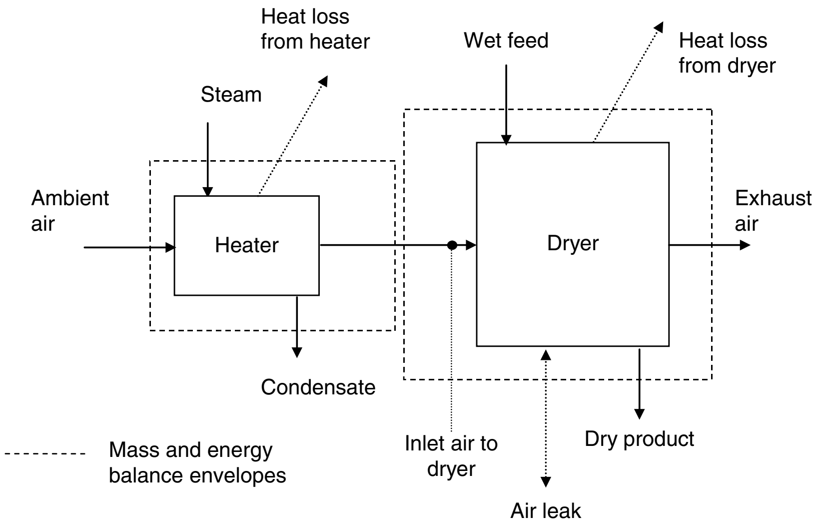


Fig. 11 Energy balances around dryer and heater.

Schemes Involving Little or No Capital Expenditure

Examples of schemes that require little or no capital expenditure include those which can be described as housekeeping measures, 1–4 below, and those that are based exclusively on a beneficial change in the dryer's operating conditions, 5–7.

1. Reducing air leaks
2. Eliminating steam leaks
3. Improving dryer insulation
4. Improving heater insulation
5. Reducing the air mass flowrate
6. Increasing the inlet air temperature
7. Eliminating over-drying

Brief outlines of each of these measures are as follows. Leakage of hot air out of the system or cold air into it clearly represent a waste of energy which can normally be reduced or eliminated by appropriate maintenance procedures. Air in-leakage through, for example, a warped access door, can occur when the dryer is fitted with an exhaust fan and is operated under a slight vacuum. It has the effect of reducing the temperature of the air in the drying chamber below the desired value. As a result, the inlet air has to be heated to a somewhat higher temperature to compensate. The primary effect of steam leaks in the heater is to increase the humidity of the inlet air to the dryer, which, in turn, may reduce the drying rate. To compensate, the inlet air temperature has to be raised above the design value in order to maintain the desired product moisture content.

Adequate insulation of the dryer and the air heater is vital in order to maintain heat losses within acceptable levels. As a result, damaged or water-laden insulation needs to be replaced as soon as possible. It is also sensible to periodically review the optimum insulation thickness, as this will naturally increase as the price of fuel escalates. Repairing and upgrading the insulation, however, will impact on the dryer's operation. Assuming that the control settings remain unchanged, the temperature of the drying air will increase as a result of the reduced heat loss and the product will therefore be over-dried. There are three possible ways in which the dryer operating conditions can be changed to overcome this problem:

1. Maintain constant solids feedrate and air flow and reduce the inlet air temperature to the dryer.
2. Maintain constant solids feedrate and inlet air temperature and reduce the air flowrate to the dryer.
3. Maintain constant inlet air temperature and flowrate and increase the solids feedrate.

Each of the above approaches needs to be evaluated in order to determine the best strategy to adopt.

It may also be possible to reduce the dryer energy consumption simply by modifying its operating conditions. Most of the heat leaves the dryer via the exhaust air stream. It can be shown that if the air flowrate is reduced, its humidity will be increased and energy will be saved. In practice, there are limits to this approach. For example, it is normally necessary to maintain a difference of at least 10°C between the temperature and dewpoint of the exhaust air in order to prevent condensation in the downstream ductwork and gas-cleaning equipment. Moreover, in a fluidized bed dryer, a minimum air velocity is required to fluidize the particles. Similar limitations exist on other types of dryer. It can also be shown that operating a dryer at as high an inlet air temperature as possible also reduces its energy consumption. Naturally, other adjustments will have to be made in order to maintain a constant product moisture content and care should be taken to avoid damage to the product if it is heat sensitive.

Based on 1978 figures for the United Kingdom, Baker and Reay[9] estimated that, on a national scale, around 11×10^6 GJ/y could be saved by implementing the low-cost measures described above. This represents about 8.6% of the energy that was then expended on drying in the United Kingdom.

Schemes Involving Significant Capital Expenditure

Only when the low-cost measures described above have been exhausted and existing dryer performance optimized should capital-intensive schemes even be considered. These possibilities include:

1. Recovering heat from the exhaust air
2. Partially recirculating the exhaust air
3. Utilizing waste heat
4. Monitoring and advanced control
5. Switching from an indirect to a direct heater
6. Prior dewatering of the dryer feedstock

Most of these techniques have been discussed by Mercer,[10] who also summarized the results of several industrial case studies. In practice, with the possible exception of monitoring and control, such schemes are rarely economically viable as retrofits because of the associated plant modifications (e.g., to the ductwork) that are required. However, when considering the purchase of a new dryer, it is worthwhile considering all the above options, which are described briefly below.

1. As noted above, most of heat leaving a dryer is contained in its exhaust air. If this can be at least partially recovered and used to preheat the incoming air to the dryer, significant energy

savings, typically 17%–40%, may result.[10] However, industrial experience has produced mixed results. Fouling of the heat exchangers by entrained particles in the dryer exhaust and, to a lesser extent, corrosion, are commonly encountered and result in poor performance and severe operating problems. These can be overcome by employing glass heat exchangers fitted with clean-in-place washing systems; however, such equipment is expensive. Other possible devices include heat wheels, heat pipes, and run-around coils.
2. An alternative means of heat recovery, which avoids the need for heat exchangers and their attendant problems, is to partially recycle, or recirculate, the exhaust air. Typically, 10%–50% of this stream is mixed with the hot inlet air from the heater before entering the dryer. Under appropriate conditions (e.g., an exhaust air temperature in excess of 120°C and a heat-resistant product), fuel savings of up to 20% can be achieved using this technique.[11]
3. In some cases, it may be possible to use process waste heat (flue gases, low-pressure steam, etc.) as a full or partial replacement for conventional energy sources. Clearly, each case must be judged on its own technical and economic merits. Principal factors to be considered include: the quantity of "free" heat available, its reliability of supply, its compatibility with the product (temperature, feasibility of direct contact, etc.), incremental cost of engineering work required to recover the heat, etc. Solar dryers,[12] which are frequently used in tropical and sub-tropical countries to dry agricultural produce in particular, represent an extreme example of the use of free energy.
4. Many older dryers are fitted with, at best, only very basic open-loop controls. Plant trials have frequently shown that effective control systems not only reduce the energy consumption of the dryer, but may also improve product quality, reduce the amount of off-spec product, and increase throughput. Effective controllers range from relatively simple devices (e.g., proportional-integral-derivative (PID) controllers) to more sophisticated types (e.g., adaptive, model-based, and fuzzy logic). The optimal choice should be based on process conditions and the type of dryer employed. Mercer[10] reported energy savings ranging from 0 to 50%, together with associated benefits of typically 0.5–1 times the direct energy savings.
5. Provided that the product being dried is compatible with the combustion gases, a direct-fired dryer should be used in preference to an indirectly heated dryer. In the former case, most (95%–98%) of the energy produced in the burner is transferred to the dryer's inlet air stream. This compares with 85% for a typical steam heater. In practice, if we also take boiler efficiency into account, a primary fuel savings of around 30% can be achieved by switching from an indirect to a direct heater.
6. Finally, given the poor thermal efficiency of many dryers, every effort should be made to minimize the moisture content of the feed by using more energy-efficient processes. In the case of liquid feedstocks, evaporation is inevitably employed to partially dewater the feedstock prior to spray drying. With solid feeds, it may be possible to use various mechanical dewatering techniques such as vacuum or compression filtration, centrifugation, etc.

Reducing Electrical Energy Consumption

To date, the emphasis in this article has been directed towards reducing thermal energy consumption. However, the electrical energy consumed by ancillary equipment such as fans, pumps, atomizers, and conveyors should not be neglected. Baker and McKenzie[8] reported that the average thermal-to-electric power ratio for the spray dryers included in their survey was around 27; this ratio is equivalent to around nine on a primary fuel basis. Therefore, consideration should always be given to implementing measures that reduce the consumption of electricity. These include installing high-efficiency electric motors, reducing the pressure drop across the system by, for example, fitting variable-speed fans rather than dampers in the ductwork, and employing bag filters rather than cyclones for dust collection.

THE FUTURE

High fuel prices and recognition of the need to cut greenhouse gas emissions are undoubtedly encouraging companies to reduce their energy consumption. In order to adapt to this new environment, dryer operators in particular will need to make some adjustments to the ways in which they purchase and operate their equipment. For example:

1. When buying a new dryer, more emphasis should be placed on lifetime cost rather than initial capital cost, which is often the case at present. This type of long term thinking can be expected to favour those options that have a lower specific energy consumption while, naturally, fulfilling all other process requirements. Contact dryers, for instance, are more efficient ($E_s \sim 2.8$ GJ/t of water evaporated) than their convective counterparts. Superheated-steam dryers are claimed to consume less than 2 GJ/t when heat recovery is employed. Other advantages include faster drying, reduced emissions, and the possibility of recovering volatile organic

compounds (VOCs). Pulsed fluidized bed and vibrofluidized bed dryers are more efficient than their conventional counterparts because they use less air. Finally, the efficiency of two-stage dryers is normally higher than that of their single-stage counterparts.
2. Despite their added cost, the heat recovery options and other energy-saving measures described above become more attractive as fuel prices rise. However, a thorough technical evaluation needs to be undertaken to ensure that any potential benefits are not negated, e.g., by fouling of the heat exchanger surfaces.
3. More effort needs to be taken to monitor the fuel consumption of individual dryers as useful energy savings can frequently be achieved by means of low-cost measures. Fitting dryers with advanced control systems can also result in quality improvements.

Collectively, these measures can help to offset rising fuel bills, combat inflation, and conserve our planet's natural resources.

CONCLUSIONS

By their very nature, dryers are major consumers of energy and are, in addition, often operated inefficiently. The techniques described in this article outline a rigorous approach to benchmarking dryer performance and suggest a hierarchy of methods for reducing their energy consumption.

REFERENCES

1. Lababidi, H.M.S.; Baker, C.G.J. Web-based expert system for food dryer selection. Comput. Chem. Eng. **2003**, *27*, 997–1009.
2. Mujumdar, A.S., Ed., *Handbook of Industrial Drying,* 2nd Ed., Marcel Dekker: New York, 1995.
3. Baker, C.G.J., Ed., *Industrial Drying of Foods*, Blackie Academic and Professional: London, 1997.
4. Richardson, J.F.; Harker, J.H.; Backhurst, J.R., 5th Ed. In *Chemical Engineering*; Butterworth Heinemann: New York, 2003, Vol. 2; 901–969 [Chapter 16].
5. Geankoplis, C.J. *Transport Processes and Unit Operations*, 3rd Ed.; Prentice-Hall: London, 1993; 520–583 [Chapter 9].
6. Baker, C.G.J. Energy efficient dryer operation—an update on developments. Drying Technol. **2005**, *23*, 2071–2087.
7. Gilmour, J.E.; Oliver, T.N.; Jay, S. Energy use for drying processes: the potential benefits of airless drying. In *Proceedings 11th International Drying Symposium*, Halkidiki, Greece, August 19–22, 1998; Vol. A, 573–580.
8. Baker, C.G.J.; McKenzie, K.A. Energy consumption of industrial spray dryers. Drying Technol. **2005**, *23*, 365–386.
9. Baker, C.G.J.; Reay, D. Energy usage for drying in selected U.K. industrial sectors. In *Proceedings 3rd International Drying Symposium*, Vol. 2, Drying Research Ltd: Birmingham, U.K., 1982, 201–209.
10. Mercer, A. Learning from experiences with industrial drying technologies. In *CADDET Energy Efficiency Analyses Series No. 12*; Sittard: Netherlands, 1994.
11. Masters, K. *Spray Drying Handbook*; Longman: London, 1985; 106–108.
12. Imre, L. Solar dryers. Baker, C.G.J., Ed. In *Industrial Drying of Foods*, Blackie Academic and Professional: London, 1997; [Chapter 10], 210–241.

Electric Motors

H. A. Ingley III
Department of Mechanical Engineering, University of Florida, Gainesville, Florida, U.S.A.

Abstract
This entry on electric motors is written from the perspective of a mechanical engineer involved in the design and specification of mechanical systems powered by electric motors. The entry describes several motor types and their intended applications. Specific components and performance criteria for motor selection are illustrated. The entry concludes with a discussion on motor efficiencies and the use of adjustable speed drives (ASDs) to enhance electric motor performance.

INTRODUCTION

This chapter on electric motors is written from the perspective of a mechanical engineer involved in the design and specification of mechanical systems powered by electric motors. The reader may have the impression that electrical engineers have the prime responsibility to specify electric motors, and it may come as a surprise that in most cases, it is the mechanical engineer specifying the electric motor. This is primarily because most of the mechanical equipment specified by mechanical engineers is packaged with electric motors. However, careful coordination with an electrical engineer on the electrical service requirements for the motor is essential in any design.

Electric motors find wide use in the industrial, commercial, and residential sectors of the United States. The following table summarizes the extent of these applications (Table 1).

Electric motors are responsible for a significant fraction of this nation's electrical consumption (Table 2). As an example, electric motor energy use represents 60% of the total energy consumption in the United States' industrial sector. A basic knowledge of electric motor design and applications equips the engineer with the skills to provide motor selections that insure reliable service with minimal energy consumption.

As might be expected, there are about as many electric motor classifications as there are applications. The criteria listed in Table 3 should be considered when selecting motors for a given application.

A general classification of common motor types with comments on their applications and special characteristics is given in Table 4. The three basic types of motors are alternating current induction/asynchronous, alternating current synchronous and direct current. Over 90% of the electrical energy input to electric motors serves to power alternating current induction motors.[2]

Because induction motors account for the highest percentage of energy use among motors, it is important to have a basic understanding of how these motors work and what technologies are available to enhance the performance of these motors. Induction motors are constructed to run on single-phase or three-phase power. Most homes in the United States are supplied with single-phase power; therefore, the majority of motors used in these homes are single-phase induction motors. Even though numerically the majority of motors fall in this category (see Table 2), the highest percentage of energy use results from the use of larger three-phase induction motors.

One other important distinction between the smaller single-phase motors and the three-phase motors is the drive method used in the application. Generally, single-phase motors have their shafts directly connected to the device; e.g., a motor driving a small hermetically sealed compressor or a small bathroom exhaust fan. When the motor fails, the whole device generally requires replacement. Larger three-phase motors are coupled to their devices with belts or couplings that allow for relatively easy motor replacement.

Another distinction between single-phase and three-phase motors is the method used to start the motor. In a three-phase motor, there are three stator (the fixed coil) windings spaced at 120° intervals around the rotor (the rotating member of the motor). These motor windings result in a rotating field around the rotor, thus starting and maintaining rotation of the rotor. Single-phase motors do not have this magnetic field arrangement; therefore, they require an auxiliary method to start the rotation. Generally, this is accomplished by a separate winding set at an angle of 90° from the main winding. This auxiliary winding is connected in series with a capacitor to provide a starting torque to initiate rotor rotation. After the motor starts, the auxiliary winding is disconnected by an internal switch. This type of motor is referred to as a capacitor-start motor.

Keywords: Electric motors; Induction motors; Slip; Motor efficiency; Power factor; Adjustable speed drives; Energy estimating.

Table 1 Electric motor application statistics

14% of the population of motors power centrifugal fans
33% are used for material handling
34% are used for positive displacement pumps
18% are used for centrifugal pumps
32.5% are used in variable torque applications
67.5% are used in constant torque applications

Source: From American Council for an Energy Efficient Economy (see Ref. 1).

In permanent split-capacitor induction motors, the capacitor is not disconnected after motor startup.

The difference between a synchronous and an asynchronous induction motor is dictated by the amount of full-load motor slip resulting from the motor design.

$$\text{Synchronous speed in rpm} = \frac{\text{applied voltage frequency(Hz)} \times 60}{\text{number of pole pairs in the motor}} \quad (1)$$

Referring to Eq. 1 above, a motor having one pole pair (two poles) would have a synchronous speed in the United States (where the applied voltage frequency is 60 Hz) of 3600 rpm. Four poles would result in a synchronous speed of 1800 rpm, and six poles would result in a synchronous speed of 1200 rpm. However, induction motors experience "slip," which results in a lower operating speed. These motors, referred to as asynchronous induction motors, can experience full-load slip in the range of 4% for small motors and 1% for large motors. Thus, a four-pole asynchronous motor would operate at 1750 rpm.

Induction motors can be further classified according to the design of their rotors. When these rotor components are formed into a cylindrical shape resembling a squirrel cage, the motor is referred to as a "squirrel cage" motor. This type of motor is relatively inexpensive and reliable; therefore, it finds wide use in commercial and industrial applications. In motors where starting current, speed, and torque require close control, the rotor is comprised of copper windings much like the stator. This requires an external source of power for the rotor, which can be accomplished with slip rings and brushes. As might be expected, this increases the first cost of the motor. Maintenance costs are also higher. These wound-rotor motors are generally sold in sizes of 20 hp and up.

Table 2 Electric motor facts

It is estimated that 60% of all electric power produced in the United States is used by electric motors
90% of all motors are under 1 hp
8% of all motors are in the 1–5 hp range
2% of all motors are over 5 hp
70% of the electric use by motors is by motors over 5 hp; 22% is by motors in the 1–5 hp range; and 8% is by motors under 1 hp

Source: From American Council for an Energy Efficient Economy (see Ref. 2).

Table 3 Electric motor selection criteria

Brake horsepower output required
Torque required for the application
Operating cycles (frequency of starts and stops)
Speed requirements
Operating orientation (horizontal, vertical, or tilted)
Direction of rotation
Endplay and thrust limitations
Ambient temperature
Environmental conditions (water, gasoline, natural gas, corrosives, dirt and dust, outdoors)
Power supply (voltage, frequency, number of phases)
Limitations on starting current
Electric utility billing rates (demand, time of day)
Potential application with variable frequency drives

When an engineer is considering an electric motor for a given application, it is important to match the enclosure of the motor to the type of operating environment involved. There are tradeoffs to be considered in this selection. A more open motor will stay cooler, which will improve its efficiency and service life. A closed motor will be less subject to contamination by a wet or dirty environment, but it may be less efficient. National Electrical Manufacturers Association (NEMA) Standard MG-1-1978 describes 20 different types of enclosures which address these two tradeoffs. Generally, in commercial and industrial applications, the enclosures fall in one of three categories: open drip proof (ODP), totally enclosed fan cooled (TEFC); and explosion proof (EXP). Fig. 1 is an example of a TEFC motor enclosure. Table 5 describes some of the other basic nomenclature used to describe motor enclosures.

Proper motor selection also includes an understanding of temperature ratings, insulation classifications, and service factors. National Electrical Manufacturers Association Standard MG 1-1998 relates the allowable temperature rises for each insulation classification. The most common insulation class is class B, which allows for a temperature rise of 80°C for Open or TEFC motors with service factors of 1.0, and 90°C for motors operating at a service factor condition of 115% rated load. When applying adjustable speed drives (ASDs) to motors, it is important to ensure that the insulation class is rated for ASD operations. In general, Class F will meet this requirement. Most premium-efficiency motors should also have this classification of insulation. At times, consideration may be given to selecting motors

Table 4 Classification of common motor types

a.c.	Induction	Squirrel cage	Three-phase, general purpose, >0.5 hp, low cost, high reliability
			Single-phase, <0.5 hp, high reliability
		Wound rotor	>20 hp, special purpose for torque and starting current regulation, higher maintenance requirement than for squirrel cage
			Very large sizes, high efficiency and reliability, higher maintenance requirement than for squirrel cage
a.c.	Synchronous	Reluctance	Standard, small motors, reliable, synchronous speed
			Switched, rugged high efficiency, good speed control, high cost
		Brushless permanent magnet (overlaps d.c. also)	High efficiency, high performance applications, high reliability
d.c.	Wound rotor	Limited reliability, relatively high maintenance requirements	Series, traction, and high torque applications
			Shunt, good speed control
			Compound, high torque with good speed control
			Separated, high performance drives; e.g., servos

Source: From American Council for an Energy Efficient Economy (see Ref. 2).

(especially large motors, 200 hp and up) with heaters in the motor windings. These heaters are energized when the motor is de-energized and they keep the motor windings warm to inhibit moisture wicking into the winding interstitial areas.

The service factor rating for a motor addresses the capacity at which an electric motor can operate for extended periods of time at overload conditions. As an example, if the service factor is 1.0, the motor cannot operate above full-load capacity for a significant period of time without damaging the motor. Service factors of 1.15, 1.25, or 1.35 indicate that a motor can operate at 1.15, 1.25, or 1.35 times its rated full load, respectively, for extended periods of time without failure. It should be noted, however, that this does not mean that the motor's service life is not affected. Insulation life can be reduced by as much as 50% operating under these conditions.

When evaluating motor replacement options, the frame size for a motor is also a matter of consideration. U-frame and T-frame designations are typical with electrical motors. New high-efficiency motors with U-frames are not always interchangeable with older style U-frame motors. It is important to check the frame size and determine if a conversion kit is required. The method of attachment or integration of the frame with the motor is also important. The author has often experienced owner preferences in selecting motors that will be used to drive centrifugal fans with a belt drive system. In one case, the owner had experienced structural failures with motors in which the motor housing was welded to the base frame.

Elec–Elec

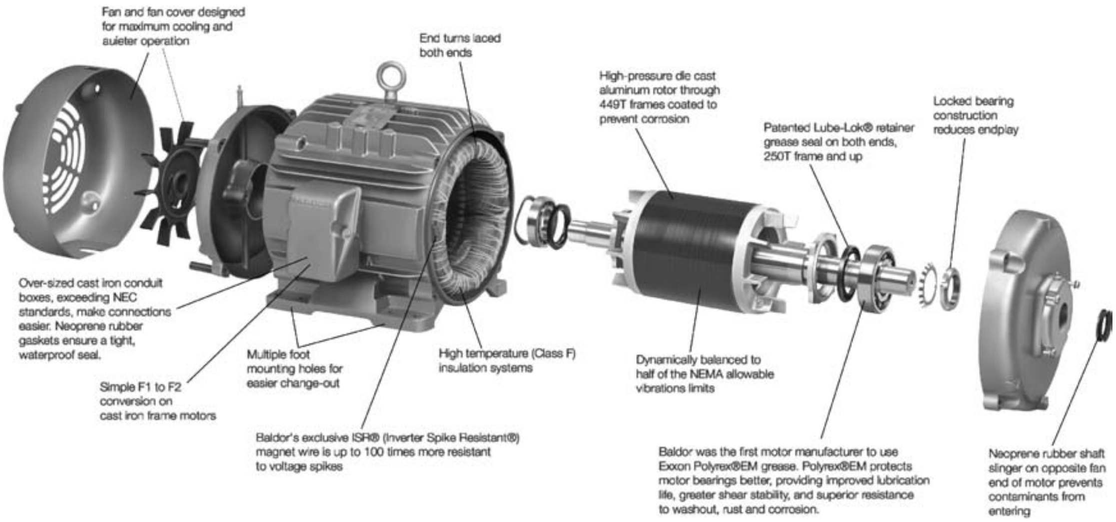


Fig. 1 Premium-efficiency motor (courtesy of Baldor).

Table 5 Examples of motor enclosure nomenclature

Open-type—full openings in frame and endbells for maximum ventilation, low cost
Semi-protected—screens on top openings to keep falling debris out, protected with screens on all openings (older style motors)
Drip-proof—upper parts are covered to keep drippings out falling at an angle not over 15° from vertical
Splash-proof—baffled at bottom to keep particles out coming from an angle of 100° from vertical
Totally enclosed—explosion proof, nonventilated, or separately ventilated for hazardous atmospheres
Fan-cooled—totally enclosed motor with fan built in to ventilate motor

This particular owner preferred motors in which the base and motor housing was a single integrated piece.

In motor replacement or in selection of new motors, the supply voltage should be coordinated with the selection. Most three-phase motors are designed to operate at 460 V and 60 Hz in the United States. Many commercial or institutional facilities are served with 208- or 230-V services. Residential motors and small fractional horsepower industrial motors would be served with 120-V power. It should be noted that even though a motor might be rated to operate at either 208 or 230 V, the operation at the lower voltage will result in a lower efficiency and shorter service life. Off-voltage operation can adversely affect electric motor performance and should be avoided.

Associated with the power supply to the motor is the motor's contribution to the overall power factor for the facility using the motor. Utilities generally penalize customers with poor power factors, because this translates into reduced availability of transformer capacity and inefficient use of power. The typical threshold level for power factors may be in the range of 85%–95% for some utilities. When specifying electric motors, the power factor should be included, as well as the other factors indicated above.

Now that we have covered some of the basics in motor types and selection, the discussion will turn to addressing three questions that consulting engineers often have to consider when evaluating options for motor replacement: what motor efficiency is most cost effective? Should motor rewinding be considered if an existing motor has failed and requires replacement? Are there other options to consider that might save energy?

The importance of motor efficiency is certainly obvious considering the energy consumption trends referenced at the introduction of this chapter. Improvements in electric motor efficiency have been recognized as a means of reducing energy consumption in all sectors of the economy.

The efficiency for motors is generally defined by Eq. 2:

$$\text{Efficiency} = \frac{746 \times \text{output horsepower}}{\text{input watts}} \qquad (2)$$

In October of 1992, the U.S. Congress signed into law the Energy Policy Act (EPAct) that established energy efficiency standards for general-purpose, three-phase alternating current industrial motors ranging in size from 1 to 200 hp. In October of 1997, the EPAct became effective.[3] Table 6 illustrates a sample of required full-load nominal efficiencies for general purpose motors. Note that the power factors are not given in this table. Consideration should be given to selecting the highest possible power factor when selecting any motor. There are several standards available for testing motors. However, for the same motor, the tests performed using these standards may result in differing efficiencies. The generally recognized standards for testing are the IEEE 112 Method B and CSA C-390-93 standards. These two testing standards typically provide the same results when applied to the same motor.

A motor that meets or exceeds the minimum efficiencies specified by the Consortium for Energy Efficiency (CEE) is referred to as a CEE premium-efficiency motor. These motors generally have efficiencies

Table 6 Required full-load nominal efficiencies for general purpose motors.

	Open motors (%)			Enclosed motors (%)		
Motor hp	**6-pole**	**4-pole**	**2-pole**	**6-pole**	**4-pole**	**2-pole**
1	80.0	82.5		80.0	82.5	75.5
5	87.5	87.5	85.5	87.5	87.5	87.5
7.5	88.5	88.5	87.5	89.5	89.5	88.5
10	90.2	89.5	88.5	89.5	89.5	89.5
15	90.2	91.0	89.5	90.2	91.0	90.2
20	91.0	91.0	90.2	90.2	91.0	90.2
100	94.1	94.1	93.0	94.1	94.5	93.6

Courtesy of Department of Energy

Table 7 Comparison of Energy Policy Act (EPAct) to premium-efficiency motors

Horsepower	Average EPAct efficiency at 75% load (%)	Average premium efficiency at 75% load (%)	Ratio of premium-efficiency motor cost to EPAct motor cost
1	82.4	85.2	1.20
5	88.5	90.5	1.20
10	91.1	91.9	1.09
20	92.3	93.5	1.08
50	93.9	94.8	1.14
75	94.5	95.7	1.09
100	94.9	95.7	1.17
150	96.1	95.9	1.24
200	95.3	96.3	1.14

Source: From American Council for an Energy Efficient Economy (see Ref. 2, Table Aa-1).

that exceed those of EPAct motors. Table 7 illustrates a comparison between EPAct motor efficiencies and CEE premium motor efficiencies for several motor sizes. Even though the price of CEE premium-efficiency motors exceeds EPAct motors by as much as 20%, energy and demand savings can result in simple paybacks of 2 to 3 years.

The EPAct of 2005 required that federal agencies select and purchase only premium efficient motors that meet a specification set by the Secretary of Energy. On August 18, 2006, the DOE set forth the specifications developed by the Federal Energy Management Program to be used for purchasing. These standards are consistent with those recommended by the NEMA and the CEE. Tables 8 and 9 illustrate these new standards. In order to meet Energy Star requirements, the efficiencies in Tables 8 and 9 must also be met.

In evaluating motors in the field for general performance; in particular, to determine if the motor is overloaded or underloaded, measurements of applied voltage and amperes can provide valuable information. This data can also be used to evaluate annual energy consumption for a motor if an estimate of operating periods can be made. Eq. 3 illustrates a method for estimating the power consumption of a motor in kilowatts[4]:

$$\text{Three-Phase Power, kW} = \frac{\text{Volts} \times \text{Amperes} \times \text{Power Factor} \times 1.732}{1000} \quad (3)$$

When replacing or rewinding motors—premium-efficiency motors (motors whose efficiencies exceed EPAct requirements)—the economics for the decision making will obviously depend on several factors such as motor type, operating regime (hours/year), and cost. Consortium for Energy Efficiency premium-efficiency motors[2] with operating regimes in the range of 4000 h per year will experience simple paybacks of 6 years or less compared to EPAct motors. Totally enclosed fan cooled motors in the size range of 2–25 hp will pay back in 3 years or less. Motors in sizes of 10 hp and below are generally replaced when they fail; larger motors are often repaired. When the rewinding or repair option of a standard-efficiency motor is compared to replacement with a premium-efficiency motor, the simple payback can be less than 2 years (for ODP motors in sizes up to 200 hp and for TEFC motors in sizes up to 40 hp). For single-phase and fractional horsepower motors, the simple payback varies significantly (2.5–10 years) when replacement with a premium-efficiency motor is evaluated.

Another area of concern consulting engineers face is in the selection of motors for pumps. Many engineers base their selection of centrifugal pump motor size on nonoverloading criteria. In order to ensure that the pump motor never overloads, the engineer checks the box on the pump selection software that will pick nonoverloading motor sizes. The resulting pump curve will always be to the left of the motor hp curve. The significance of this is that a pump with a duty horsepower of 2.3 hp could have a 5-hp motor selected. This results in the motor operating many hours of the year at part load and consequently, at lower than full-load efficiencies. Consulting engineers should evaluate their decisions to select nonoverloading motors with energy efficiency in mind. With the advent of reliable electronic ASDs, significant energy savings are now available for many applications involving electrically driven pumps and fans. It is very common today to design commercial building air conditioning systems with ASD technology. Variable-air-volume air conditioning systems use ASDs to vary fan speed as the demand for supply air varies with cooling and heating loads. Similarly, in large chilled-water or heating-hot-water distribution systems, ASDs are used to vary the flow rate of water in response to varying loads. In one such application involving two 400-hp circulating pumps on a chilled water system, the

Table 8 Nominal efficiencies for induction motors rated 600 V or less

	Random wound					
	Open drip-proof (%)			Totally enclosed fan-cooled (%)		
Horsepower	6-pole	4-pole	2-pole	6-pole	4-pole	2-pole
1	82.5	85.5	77.0	82.5	85.5	77.0
1.5	86.5	86.5	84.0	87.5	86.5	84.0
2	87.5	86.5	85.5	88.5	86.5	85.5
3	88.5	89.5	85.5	89.5	89.5	86.5
5	89.5	89.5	86.5	89.5	89.5	88.5
7.5	90.2	91.0	88.5	91.0	91.7	89.5
10	91.7	91.7	89.5	91.0	91.7	90.2
15	91.7	93.0	90.2	91.7	92.4	91.0
20	92.4	93.0	91.0	91.7	93.0	91.0
25	93.0	93.6	91.7	93.0	93.6	91.7
30	93.6	94.1	91.7	93.0	93.6	91.7
40	94.1	94.1	92.4	94.1	94.1	92.4
50	94.1	94.5	93.0	94.1	94.5	93.0
60	94.5	95.0	93.6	94.5	95.0	93.6
75	94.5	95.0	93.6	94.5	95.4	93.6
100	95.0	95.4	93.6	95.0	95.4	94.1
125	95.0	95.4	94.1	95.0	95.4	95.0
150	95.4	95.8	94.1	95.8	95.8	95.0
200	95.4	95.8	95.0	95.8	96.2	95.4
250	95.4	95.8	95.0	95.8	96.2	95.8
300	95.4	95.8	95.4	95.8	96.2	95.8
350	95.4	95.8	95.4	95.8	96.2	95.8
400	95.8	95.8	95.8	95.8	96.2	95.8
450	96.2	96.2	95.8	95.8	96.2	95.8
500	96.2	96.2	95.8	95.8	96.2	95.8

Table 9 Nominal efficiencies for induction motors rated 5 kV or less

	Form wound					
	Open drip-proof (%)			Totally enclosed fan-cooled (%)		
Horsepower	6-pole	4-pole	2-pole	6-pole	4-pole	2-pole
250	95.0	95.0	94.5	95.0	95.0	95.0
300	95.0	95.0	94.5	95.0	95.0	95.0
350	95.0	95.0	94.5	95.0	95.0	95.0
400	95.0	95.0	94.5	95.0	95.0	95.0
450	95.0	95.0	94.5	95.0	95.0	95.0
500	95.0	95.0	94.5	95.0	95.0	95.0

author observed a payback of under 3 years for a major retrofit to variable speed pumping. This project also included modifications to the control valves to accommodate the variable flow strategy.

The application of ASD technology does require care. Consider the effect on power factor when selecting an ASD. Some ASD applications can actually result in an installed system with a power factor better than the original motor power factor.

However, improper selection of an ASD can also result in a lower overall power factor. Another concern deals with the harmonics that an ASD can superimpose on the system. From a consulting mechanical engineer's perspective, this is an electrical engineering problem! Consulting with the project's electrical engineer on the effects of harmonics is important. Applying too many ASDs on an electrical system can cause harmful harmonics that can affect the overall system performance. Many ASD suppliers offer a service for evaluating these harmonics and can assist in dealing with this issue.

Another significant issue in the selection of an ASD for an application is having a reasonable estimate of the operating profile for the system.[1] Without this knowledge, energy savings associated with the ASD application can be over- or even underestimated. Once this issue is resolved, and assuming the estimate is favorable, the resulting energy savings are still in question if the ASD system is not maintained properly. The total success of the ASD system hinges on the accuracy and reliability of sensing pressure differentials or flow rates. If the sensor goes out of calibration or ceases to function, the ASD will be rendered useless.

Because of the complex electronics in an ASD system (at least, complex to the consulting mechanical engineer!) and the various sensors and controllers that might be connected to the ASD, lightning and surge protection is another consideration in the overall system design. Again, the consulting electrical engineer should be able to provide insight into the appropriate lightning- and surge-protection systems for the project. The use of fiber optics in the communication system between the sensors and the ASD can eliminate some of the problems from lightning. Surge protection and optical relays can provide additional protection.

One last consideration in the control of electrical motors addresses the method of starting electric motors. There are many technologies available that can be used to start motors (such as mechanical motor starters, ASDs, and soft-start starters) and improve the life expectancy of these motors. The selection of the motor starter sometimes falls on the mechanical engineer, and sometimes the electrical engineer. Whichever the case, consideration should be given to coordinating the starter selection with the motor selection to ensure the most efficient combination results. The electrical engineer's role at this point is also significant. Careful coordination is required to verify and make necessary adjustments to the electrical service for the motor, to verify branch circuit wiring capacity, and to coordinate branch circuit protection as well as overload protection and disconnect requirements. Then it would be prudent to check the motor rotation (or else pumps will run backwards!) and alignment.

CONCLUSION

In summary, electric motor selection, replacement, and repairs require careful evaluation in order to provide an efficient, reliable, and easy-to-maintain motor drive system whether in the residential, commercial, or industrial sector. With over 60% of our nation's industrial energy use represented by the use of electric motors, it is irresponsible not to apply the utmost of care in our practice of motor selection and replacement.

REFERENCES

1. Elliott, R.N.; Nadel, S. *Realizing energy efficiency opportunities in industrial fan and pump systems*, Report A034; American Council for an Energy Efficient Economy: Washington, DC, April 2003.
2. Nadel, S.; Elliott, R.N.; Shepard, M.; Greenberg, S.; Katz, G.; de Almeida, A.T. *Energy-Efficient Motor Systems: A Handbook on Technology, Program, and Policy Opportunities*, 2nd Ed.; American Council for an Energy Efficient Economy: Washington, DC, 2002.
3. *Engineering Cookbook*, 2nd Ed.; Loren Cook Company: Springfield, MO, 1999.
4. Turner, W.C., Ed. *Energy Management Handbook* 5th Ed.; Fairmont Press: Atlanta, GA, 2005.

Electric Power Transmission Systems

Jack Casazza
American Education Institute, Springfield, Virginia, U.S.A.

Abstract
This entry discusses electric power transmission system functions, the benefits produced, components, ratings and capacity, alternating current (AC) and direct current (DC) transmission, transmission system operations, the need for coordination, control areas, NERC and reliability councils, reliability standards, transmission access, and new technology.

TRANSMISSION FUNCTIONS

Basic Role

The transmission system provides the means by which large amounts of power are delivered from the generating stations where it is produced to other companies or to locations where voltage is reduced, to supply subtransmission systems or substations where it is distributed to consumers. Transmission in the United States is mostly by three phase, 60 Hz (cycles per seconds) alternating current (AC) at voltages between 115,000 and 765,000 V. Direct current (DC) is used at a few locations but its potential role is increasing.

Benefits of Interconnection[1]

Electric power must be produced at the instant it is used. Needed supplies cannot be produced in advance and stored for future use. It was soon recognized that peak use for one system often occurred at a different time than peak use in other systems. It was also recognized that equipment failures occurred at different times in various systems. Engineering analyses showed significant economic benefits from interconnecting systems to provide mutual assistance. The investment required for generating capacity could be reduced. Reliability could be improved. Differences in the cost of producing electricity in the individual companies and regions often resulted in one company or geographic area producing some of the electric power sold to another company in another area for distribution. This lead to the development of local, then regional, and subsequently, five grids in North America with three United States transmission grids, as shown in Fig. 1. Fig. 2 shows the key stages of the evolution of these transmission grids.

Keywords: Electric power; Transmission systems; Interconnections; Voltage control; Synchronous operation; Capacity; Losses; Reliability.

Summarizing the transmission system[2]:

- Delivers electric power from generating plants to large consumers and distribution systems.
- Interconnects systems and generating plants to reduce overall required generating capacity requirements by taking advantage of:
  - The diversity of generator outages, i.e., when outages of units occur in one plant, units in another plant can provide an alternative supply.
  - The diversity of peak loads because peak loads occur at different times in different systems.
- Minimizes fuel costs in the production of electricity by allowing its production at all times at the available sources having the lowest incremental production costs.
- Facilitates the location and use of the lowest cost additional generating units available.
- Makes possible the buying and selling of electric energy and capacity in the marketplace.
- Helps provide for major emergencies such as hurricanes, tornadoes, floods, strikes, fuel supply disruptions, etc.

Transmission of Real and Reactive Power[3]

The delivery of electric energy to perform desired functions requires the delivery of “real” and “reactive power.” The real power is produced in the power plant from fuels or energy sources and provides the energy that is used. This real power must be accompanied by reactive power that provides the electric fields required by various devices for the utilization of this energy. This reactive power does not include any energy. It is produced by the fields of the generators, by capacitors installed for that purpose, and by the “charging current” of the transmission system.

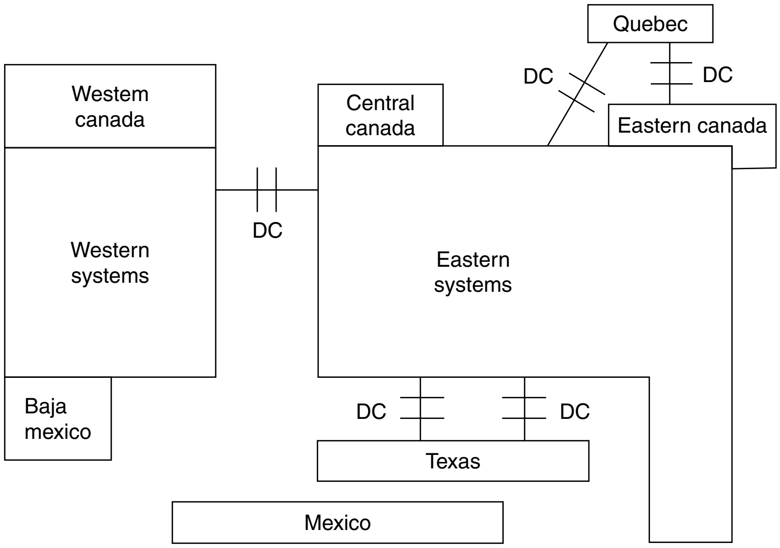


Fig. 1 The five synchronous systems of North America.

Electrical Characteristics[4]

Transmission systems have resistance (R), which causes the heating of conductors and the loss of energy when current (I) flows and reactance (X), which causes voltage drops when current flows. Reactance can be positive or negative; positive when it is an inductive reactance and negative when it is a capacitive reactance.

TRANSMISSION COMPONENTS

Transmission Lines

The transmission system consists of three-phase transmission lines and their terminals, called substations or switching stations. Transmission lines can be either overhead or underground (cable). High-voltage alternating current (HVAC) lines predominate, with high-voltage direct current lines (HVDC) used for special applications. Overhead transmission, subtransmission, and primary distribution lines are strung between towers or poles. In urban settings, underground cables are used primarily because of the impracticality of running overhead lines along city streets. While underground cables are more reliable than overhead lines (because they have less exposure to climatological conditions such as hurricanes, ice storms, tornadoes, etc.), they are also much more expensive than overhead lines to construct per unit of capacity and they take much longer to repair because of the difficulty in finding the location of a cable failure and replacement.

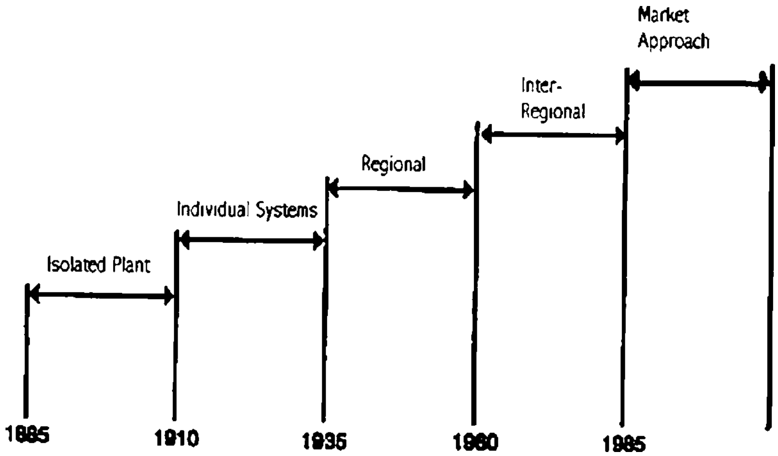


Fig. 2. Stages of transmission system development.

The primary components of an overhead transmission line are:

- Conductors (three, one per phase)
- Ground or shield wires
- Insulators
- Support structures
- Land or right-of-way (ROW)

Conductors consist of stranded aluminum woven around a core of stranded steel that provides structural strength. When there are two or more of these wires per phase, they are called bundled conductors.

Ground or shield wires are wires strung from the top of one transmission tower to the next, over the transmission line. Their function is to shield the transmission line from lightning strokes. Insulators are used to attach the energized conductors to the supporting structures, which are grounded. The higher the voltage at which the line operates, the longer the insulator strings.

The most common form of support structure for transmission lines is a steel lattice tower, although wood H frames (so named because of their shape) are also used. In recent years, as concern about the visual impact of these structures has increased, tubular steel poles also have come into use. The primary purpose of the support structure is to maintain the electricity carrying conductors at a safe distance from the ground and from each other. Higher-voltage transmission lines require greater distances between phases and from the conductors to the ground than lower-voltage lines and therefore they require bigger towers. There has been some concern about the biological effects of transmission lines with the general conclusion being that there are no serious effects.[5]

Ratings

The capability of an individual overhead transmission line or its rating is usually determined by the requirement that the line does not exceed code clearances with the ground. As power flows through the transmission line, heat is produced. This heat will cause an expansion of the metal in the conductor, and as a result increase the amount of its sag. The amount of sag will also be impacted by the ambient temperature, wind speed, and sunlight conditions. The heating can also affect the characteristics of the metals in the conductors, reducing their strength.

Ratings are usually of two types—normal and emergency—and are usually determined for both summer and winter conditions. Some companies average the summer and winter ratings for the fall and spring. Ratings are also specified for various time periods. A normal rating is the level of power flow that the line can carry continuously. An emergency rating is the level of power flow the line can carry for various periods of time, for example, 15 and 30 min, 2, 4, and 24 h, and so forth.

In recent years, there has been a trend in calculating ratings for critical transmission lines on a real-time basis, reflecting actual ambient temperatures as well as the recent loading (and therefore heating) patterns.

Cables

The majority of the transmission cable systems in the United States are high-pressure fluid filled (HPFF) or high-pressure liquid filled (HPLF) pipe-type cable systems. Each phase of a high-voltage power cable usually consists of stranded copper wire with oil-impregnated paper insulation. All three phases are enclosed in a steel pipe. The insulation is maintained by constantly applying a hydraulic pressure through an external oil adjustment tank to compensate for any expansion or shrinkage of the cable caused by temperature variations.

Cable capacity is determined by the effect of heat on the cable insulation. Because the cable is in a pipe that is buried in a trench, dissipation of the heat is a major issue in cable design and operation. Cable capacity can be increased by surrounding the pipe with thermal sand, which helps dissipate heat.

A limitation on the application of ac cables is their high capacitance, which increases with their length. Cable capacitance causes a charging current to flow equal to the voltage divided by the capacitive reactance. This can limit the length of cable that can be used without some intermediate location where shunt reactor compensation can be installed to absorb the charging current.

Substations

Substations are locations where transmission lines are tied together. They fulfill a number of functions:

- Allow power from different generating stations to be fed into the main transmission grid.
- Provide for interconnections with other systems.
- Provide transformers to be connected to feed power into the subtransmission or distribution systems. Transformers are generally equipped to change voltage ratios, with fixed taps that require de-energization to change and tap changing equipment that can change taps while the transformer is in operation.
- Allow transmission lines to be independently switched to isolate faulty circuits or for maintenance.
- Provide a location where compensation devices such as shunt or series reactors or capacitors can be connected to the transmission system.
- Provide a location for protection, control, and metering equipment.

Substation equipment includes:

- Bus work through which lines, transformers, etc., are connected.
- Protective relays that monitor voltages and currents and initiate disconnection of lines and equipment in the event of failures or malfunctions.
- Circuit breakers that interrupt the flow of electricity to de-energize facilities.
- Shunt capacitors to help provide needed reactive power.
- Disconnect switches.
- Lightning arrestors.
- Metering equipment.
- System control and data acquisition (SCADA) systems.
- Shunt reactors to limit high voltages.
- Series reactors to increase the impedance of lines.
- Phase angle regulating transformers and other devices to control power flow and voltage in specific circuits.

The bus/circuit breaker connection arrangements used in substations affect substation costs, ease of maintenance, and transmission reliability.[6]

Direct Current Transmission

An alternate means of transmitting electricity is to use HVDC technology. Direct current facilities are connected to HVAC systems by means of rectifiers, which convert alternating current to direct current, and inverters, which convert direct current to alternating current.

Starting in the 1970s, thyristors became the valve type of choice for the rectifiers and inverters. Thyristors are controllable semiconductors that can carry very high currents and can operate at very high voltages. They are connected in series to form a valve, which allows electricity to flow during the positive half of the alternating current voltage cycle but not during the negative half. Because all three phases of the HVAC system are connected to the valves, the resultant voltage is unidirectional but with some residual oscillation. Smoothing reactors are provided to dampen this oscillation.

High-voltage direct current lines transmission lines can either be single pole or bipolar, although most are bipolar—they use two conductors operating at different polarities, such as +/−500 kV. There have been a number of applications for DC transmission:

- To transmit large amounts of power over long distance is not feasible with AC.
- To transmit power across water where the capacitance of AC cables would limit circuit capacity.
- The need to connect two AC systems in a manner that prevents malfunctions or failures in one system from causing problems in another system.
- To provide direct control of power flow in a circuit.
- To limit short-circuit duties.
- To increase the ability to transfer power over existing right-of-ways because DC requires two conductors versus three for AC.

The difficulty of DC is its higher costs and the lack of reliable DC current breakers.

HOW TRANSMISSION SYSTEMS WORK

Synchronous Operation

Because of the synchronous operation of all generators, an interconnected electric power system functions as a single large machine that can extend over thousands of miles. One of this machine's characteristics is that changes in any one portion of it instantly affect the operation of all other portions. Future plans for any one part of it can affect transmission conditions in other parts of it.

In system operation, the effects of contingencies in one system can be felt throughout a large geographic area. For example, if a large generating unit is lost in New York city, the interconnected power system (the machine) suddenly has less power input than power output and slows down. As the individual rotors of every generator in this system slow down in unison and system frequency declines, each rotor gives up a certain amount of its rotating energy (ω^2R) to compensate for the lost input from the unit that has tripped off.

Instantaneously, with the loss of a large unit or power plant, there is an inrush of power from units all over the synchronous network, feeding into the system or region that has lost the generator unit. While these inrushes of power from individual units long distances away are not large, they accumulate and build up like water flowing from creeks into a river, increasingly loading the transmission lines near the system that has lost generation. This surge of power exists until automatic generator controls cause them to increase their output to compensate for the lost generating capacity and restore system frequency.

The very reason that systems have been tied together and operate in synchronism causes this effect. By having the various generator units throughout the region assist with the loss of a large generator unit in a specific system, the total amount of spare or reserve generating capacity required can be reduced. This is similar to the insurance business. The larger the number of policy holders, the better able the insurance company is to cope with any specific major disaster and the less percentage reserves it requires. The great strength of operating in synchronism and being tied together in an integrated system is the ability of one system to be helped by others. Its greatest weakness, however, is that the inrush of power into any one system can cause transmission system overloads in other systems.

Because of these characteristics of modern electric power systems, the design and operation of the key elements in a synchronous network must be coordinated. Business decisions, government legislation and regulations, and other institutional processes must be compatible with the technical characteristics. Many problems can be solved by technical solutions, some can be solved by institutional solutions, and in some cases, problems can be solved by coordinating both.

Kirchhoff's Laws[7]

In AC transmission networks, the flow of power in the various circuits is determined by Kirchhoff's Laws. Power flow is determined by the impedance of the individual circuits, the location in the network of the sources of power, and the location of the substations to which power is delivered. The flow of electrical power in the synchronous grid does not respect company boundaries, contracts or ownership of facilities. Power cannot be scheduled to flow over a specific line or lines, but will divide in accordance with Kirchhoff's Laws.

This sometimes results in two phenomena known as "parallel path flow" and "loop flow." Parallel path flow results when power to be delivered from system A to system B goes through systems in parallel. Such flows can increase transmission loadings in these other systems, reducing their ability to be used by the owners for their own purposes. Loop flows are circulating flows that occur with all systems supplying their own loads from their own sources. They are the result of network characteristics and often the result of deliberate network designs to limit total transmission investment requirements.

Transmission Capacity

Transmission limits can be determined by a number of factors, a common one being the maximum allowable thermal loading on circuits. Current flowing through conductors causes heating, which causes expansion and "sagging," reducing clearance to ground, possibly contacting trees or other obstacles, and resulting in a circuit trip out.

Potential stability disturbances can limit the amount of power that can be transferred. If stability limits are exceeded, the occurrence of a critical fault can cause generators to oscillate wildly and trip out. Voltage conditions can limit the ability to transfer power. An inadequate supply of reactive power can cause transmission voltages to become too low, causing excessive transmission currents and voltage instability, resulting in circuit trip outs.

While recognizing these various causes of transmission limits, it is also essential to recognize that the ability to deliver power is the ability of the interconnected network, i.e., the system that forms the grid, to transfer power, not the sum of the capacities of the individual circuits involved. Often, one circuit can become overloaded while another has unused capacity because of Kirchoff's Laws.

Transmission Access

The restructuring and deregulation of the electric power industry has lead to significant changes in the use of transmission systems. The owners of transmission lines must make them available to everyone who wants to use them on the same basis as they are used for their own customers. This has increased the complexity and difficulty of planning and operating transmissions systems because there are many more potential users whose decisions affect the entire system.[8]

Short Circuit Duties

An often overlooked factor is that transmission systems must have the ability to interrupt the very high currents that result when short circuits occur. This is done by circuit breakers that must have an interrupting capacity sufficient to interrupt the magnitudes of the fault currents involved. These fault currents are increased as generation and new transmission lines are added to the system. Transmission system design requires studies of the ability of circuit breakers to interrupt expected faults. When breaker capability is not adequate, expensive replacements or changes in substation or system designs may be needed.

Transmission System Losses

Power losses in the transmission system consume 3%–5% of the electric power produced. Because they are supplied by the highest cost generation available, they are responsible for more than 5% of the cost of energy produced. Many believe that the deregulation of the electric power industry has increased transmission losses.

There are two basic types of power (MW) losses in transmission systems:

- Core losses are dissipated in the steel cores of transformers. These losses typically vary with at least the third, and often higher, powers of the voltage variations at their terminals. If transmission voltages are fairly constant, core losses are also constant. An increase in the power carried by a transmission system does not substantially affect the core losses, but a variation in transmission voltage can.
- Conductor losses are dissipated in transmission lines and cables and transformer windings. As these losses depend on the resistance of the circuit (R) and vary with the square of the current and therefore

Elec–Elec

approximately with the square of the power carried by each component, they vary greatly between light load and heavy-load times and are also affected by increases in the power carried by the transmission system.

There are also reactive power (MVAR) losses in the transmission system. These losses depend on the reactance of the system (X) and again vary as the square of the current.

When electrical energy is transported across large distances through the transmission system, a portion of the energy is "lost." For a given amount of electric power, the higher the operating voltage of the transmission line, the lower the current flowing through the line. Therefore, use of higher transmission voltages permits the transmission of electric power with lower currents and a resulting reduction in energy losses.

The losses in the transmission system at a specific moment, typically at the time of system peak, are measured in megawatts and are referred to as "capacity losses." The energy expended in transmission losses over a given period is given in megawatt-hours and is referred to as "energy loss." Capacity losses require the installation of additional generation and transmission equipment; energy losses require the consumption of fuel or equivalent energy sources. The increase in a system's total losses due to a specific action is referred to as an "incremental loss."

In addition to real power losses, significant reactive power losses occur in the transmission systems, and these reactive losses are typically about 4 or 5 times the real power capacity losses.

OPERATION OF ELECTRIC BULK POWER SYSTEMS[9,10]

Need for Coordination[11,12]

The operation of the bulk power system in the United States involves the interdependency of the various entities involved in supplying electricity to the ultimate consumers.[13] These interdependencies have evolved as the utility industry grew and expanded over the century.

In a power system, the coordination of all elements of the system and all participants are required from an economic and a reliability perspective. Generation, transmission, and distribution facilities must function as a coordinated whole. The scheduling of generation must recognize the capability of the transmission system. Voltage control must involve the coordination of reactive power supplies from all sources, including generators, the transmission system, and distribution facilities. Actions and decisions by one participant, including decisions not to act, affect all participants.

In parallel with its early growth, the industry recognized it was essential that operations and planning of the system be coordinated and organizations were formed to facilitate the joint operation and planning of the nation's electric grid. Initially, holding companies and then power pools were established to coordinate the operation of groups of companies. Recently, new organizations, independent system operators (ISOs) and regional transmission operators (RTOs), have been formed to provide this coordination.

Reliability Councils and Nerc[14]

After the Northeast Blackout of 1965, regional electric reliability councils were formed to promote the reliability and efficiency of the interconnected power systems within their geographic areas. These regional councils joined together shortly afterwards to form a national umbrella group, NERC—the North American electricity reliability council. At present, there are ten regional councils. Each Council has a security coordinator who oversees the operation of the grid in their region.

The members of NERC and these regional councils come from all segments of the electric industry; investor-owned utilities; federal power agencies; rural electric cooperatives; state, municipal and provincial utilities; independent power producers; power marketers; and end-use customers. These entities account for virtually all the electricity supplied in the United States, Canada, and a portion of Baja California North, Mexico.

When formed in 1968, the NERC operated as a voluntary organization to promote bulk electric system reliability and security—one that was dependent on reciprocity, peer pressure, and the mutual self-interest of all those involved.

The growth of competition and the structural changes taking place in the industry have significantly altered the incentives and responsibilities of market participants to the point that a system of voluntary compliance is no longer adequate. New federal legislation in the United States has required formation of an electric reliability organization (ERO) to monitor and enforce national reliability standards under FERC oversight. In response to these changes, NERC is transforming itself into an industry-led self-regulatory reliability organization ERO that will develop and enforce reliability standards for the North American bulk electric system.

Control Areas

While overall system control is, in some cases, the responsibility of newly formed ISOs and RTOs, more than 140 "control areas" still perform needed functions.

A control area can consist of a generator or group of generators, an individual company, or a portion of a company or a group of companies providing it meets certain certification criteria specified by NERC. It may be a specific geographic area with set boundaries or it may be scattered generation and load.

The control centers require real-time information about the status of the system. This information includes power line flows, substation voltages, the output of all generators, the status of all transmission lines and substation breakers (in-service or out-of-service), and transformer tap settings. Some areas are implementing real-time transmission line rating systems requiring additional information such as weather conditions, conductor temperatures, and so forth.

Each control area monitors on an on-going basis the power flow on all of its interties (in some cases delivery points) and the output of each generator within its control. The sum of the internal generation and the net flow on the interties is equal to the consumer load and all transmission losses within the area.

The various commercial interests that are involved within the area are required to notify the control area personnel of their contractual arrangements on an ongoing basis for either sales or purchases of electricity with entities outside the area's boundaries.

Oasis and Transmission Capacity

The open access same-time information system (OASIS) is an Internet-based bulletin board that gives energy marketers, utilities, and other wholesale energy customers real-time access to information regarding the availability of transmission capacity. OASIS provides the ability to schedule firm and nonfirm transactions.

The North American electricity reliability council[15] has defined transmission capacity as follows:

Available transfer capacity (ATC) = Total transfer capability (TTC) − Existing commitments − transmission reliability margin (TRM) − Capacity benefit margin (CBM),

where:

- Available transfer capability is a measure of the transfer capability remaining in the physical transmission network for further commercial activity over and above already committed uses.
- Total transfer capability is the amount of electric power that can be transferred over the interconnected transmission network in a reliable manner while meeting all of a specific set of defined pre- and post-contingency system conditions.
- Transmission reliability margin is the amount of transmission transfer capability necessary to ensure that the interconnected transmission network is secure under a reasonable range of uncertainties in system conditions.
- Capacity benefit margin is the amount of transmission transfer capability reserved by load serving entities to ensure access to generation from interconnected systems to meet generation requirements in emergencies.

With this information, the control area operators can compare the total scheduled interchange into or out of the control area with the actual interchange. If the receipt of electricity exceeds the schedule, the control area must increase generation levels. If the receipt is too low, generation within the control area is reduced. These schedules are typically made a day ahead and then adjusted in real time. Because these adjustments are ongoing simultaneously by all control areas, the adjustments balance out.

The process where individual contracts scheduled within OASIS are identified as to source and customer is known as tagging. This information, while it may be commercially sensitive, is critical if system operators are to adjust system power flows to maintain reliable levels.

Concurrently, the system operators can also evaluate the expected power flows internal to the control area to determine if adjustments are required in the generation pattern to insure that all internal transmission facilities are operated within the capabilities.

Each control area also participates in maintaining the average system frequency at 60 Hz. The system frequency can deviate from normal when a large generating unit or block of load is lost. In addition to adjustments made because of variations of tie flows from schedule, another adjustment is made to correct frequency deviations.

Reliability standards have been developed by the regional reliability councils and NERC for many years. They define the reliability aspect of the interconnected bulk electric systems in two dimensions:

- Adequacy—the ability of the electric systems to supply the aggregate electrical demand and energy requirements of their customers at all times, taking into account scheduled and reasonably expected unscheduled outages of system elements.
- Security—the ability of the electric systems to withstand sudden disturbances such as electric short circuits or unanticipated loss of system elements.

Detailed reliability standards exist that specify allowable system voltage and loading conditions for various system contingencies. These are developed by the various regional reliability councils and must meet the minimum standards established by NERC. There are standards for various single contingencies and for various combinations of outages.

Elec–Elec

Meeting reliability standards in the planning of the transmission system is difficult because the time required to install new transmission is longer than the time required to install new generation. Attempting to meet reliability standards in planning for future transmission needs involves considerable uncertainties because future generation locations are not known. The general industry consensus is that the restructuring and deregulation of the electric power industry has resulted in a decrease in reliability. How will this affect future transmission policies is uncertain.

NEW TECHNOLOGIES

Future developments may have long-range effects on transmission requirements, transmission system characteristics, and the capacity of the various networks. New transmission technologies, some involving "power electronics," are under study. These include:

- The development of methods to control the division of flow of power in AC networks.
- The development of "smart systems" or "self-healing" systems[16] that may involve a redesign and upgrade of presently electromechanically controlled transmission systems.
- The subdivision of huge synchronous AC networks into smaller synchronous networks interconnected by DC.

There are other possible developments that may have a significant effect on transmission systems, including:

- The development of significant amounts of small distributed generation, including the use of solar energy, wind power, micro turbines, etc.
- A major national shift to large nuclear or coal units to reduce dependence on foreign oil and gas.
- The development of low-cost energy storage devices to allow power to be produced at one time for use at another.
- Increasing use of hydrogen as a mechanism for transferring energy from one location to another, including possible linking of hydrogen production with off peak generating capacity.

The future holds many uncertainties and requires analyses similar to post national power surveys[17] to determine how to develop transmission systems to meet potential future developments. Failure to make such analyses will result in wasteful transmission additions and design of a poor system.

REFERENCES

1. Casazza, J.A. The development of electric power transmission—the role played by technology, institutions, and people, *IEEE Case Histories of Achievement in Science and Technology*; Institute of Electrical and Electronic Engineers: New York, 1993, www.lulu.com.
2. Casazza, J.A.; Delea, F. *Understanding Electric Power Systems: An Overview of the Technology and the Marketplace*; Wiley: New York, 2003; 17.
3. Casazza, J.A.; Delea, F. *Understanding Electric Power Systems: An Overview of the Technology and the Marketplace*; Wiley: New York, 2003; 34, 35.
4. Casazza and Delea *Understanding Electric Power Systems: An Overview of the Technology and the Marketplace*; Wiley: New York, 2003; 25.
5. *Biological Effects of Power Frequency Electric and Magnetic Fields*, U.S. Congress Office of Technology Assessment, 1989.
6. Casazza, J.A.; Delea, F. *Understanding Electric Power Systems: An Overview of the Technology and the Marketplace*; Wiley: New York, 2003; 81.
7. Named after Gustav Robert Kirchoff, a German physicist, (1824–1887).
8. Non-Technical Impediments to Power Transfers, National Regulatory Research Institute, 1987.
9. Nagel, T.J. Operation of a major electric utility today. Science **1978**, *2* (4360).
10. Rustebakke, H.M. *Electric Utility Systems and Practices*, 4th Ed.; Wiley: New York, 1983.
11. Casazza, J.A. *Coordinated Regional EHV Planning in the Middle Atlantic States—U.S.A.*; CIGRE paper No. 315: Paris, France, 1964.
12. Casazza, J.A. The Development of Electric Power Transmission—The Role Played by Technology, Institutions, and People. IEEE Case Histories of Achievement in Science and Technology, Institute of Electrical and Electronic Engineers, www.lulu.com, 1993; 81.
13. Hughes, T.P. Systems builders technology's master craftsmen. CIGRE Electra **1986**, *109*, 21–30.
14. See www.NERC.com.
15. See www.NERC.com.
16. Stahlkoph, K.; Sharp, P.R. Where technology and politics meet. Electric Power Transmission Under Deregulation, IEEE, 2000.
17. Federal power commission *National Power Survey—A Report by the Federal Power Commission*; U.S. government printing office: Washington, DC, 1964.

Electric Power Transmission Systems: Asymmetric Operation

Richard J. Marceau
University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Abdou-R. Sana
Montreal, Québec, Canada

Donald T. McGillis
Pointe Claire, Québec, Canada

Abstract

Asymmetric operation of an electric power transmission corridor is an operating strategy that enables a 3-phase line to be operated with one or two phases out of service in the case of single-line transmission corridors, or with one, two, or three phases out of service in the case of multiple-line corridors, while preserving 3-phase symmetrical operation at corridor extremities. This article shows how asymmetric operation is implemented, how much it costs, and how it can improve the reliability and economics of electric power transmission systems.

Elec–Elec

INTRODUCTION

Existing electric power transmission systems are operated with three physically different systems of conductors referred to as "phases," where the sinusoidal voltages and currents in each phase are offset with respect to one another to take advantage of Nicolas Tesla's groundbreaking invention of the 3-phase alternating-current (AC) generator and motor. The 3-phase power generator is a remarkably robust and economical technology that creates the three sets of voltages and currents, the so-called phases, in 3-phase transmission lines. This results in a steady torque characteristic, which in turn translates into high generator reliability. The 3-phase AC motor, which requires an infeed of the three different phases to operate, is also a highly reliable and economic piece of machinery, and is a workhorse of industry.

Because 3-phase generators and motors are so widely used, they are connected to 3-phase high-voltage (HV), extra-high-voltage (EHV), or ultra-high-voltage (UHV) power transmission systems that are symmetrically operated, meaning that if any problem develops on one or more phases, all three phases are taken out of service. This requirement is due to the fact that if symmetric operation is not enforced, undesirable voltages and currents are generated that are harmful not only to rotating machinery, but also to many other types of loads. There are exceptions to this rule: For economical reasons, single-phase systems equipped with appropriate mitigating measures have been used in railway electrification and for distributing power at the household level. Throughout the world, however, standard practice is to operate bulk 3-phase power transmission systems symmetrically.

Symmetric Operation of Power Transmission Systems

When at least one phase of a power transmission line touches an object, causing the current of that phase to be redirected either to the ground (i.e., through a tree) or to another phase, such short-circuit conditions are generally referred to as "faults." Under normal operation, when a fault condition occurs, circuit breakers at both ends of the line interrupt the current in all three phases, effectively removing the line from operation until such time as the fault condition has been removed, even though one or two healthy phases remain that conceivably could carry useful power. Such a strategy has several disadvantages. First, a problem on 33% of a line automatically deprives the network of 100% of this same line. Second, the operation of three phases as a single organic whole rather than three independent conductors augments the probability of loss of 3-phase transmission by nearly a factor of three with respect to single-phase transmission. Finally, after a fault is cleared, the power system operator must apply remedial measures, such as changing the amount of generation from different power plants, to redirect power flows and ensure that the system is capable of sustaining further contingencies with no impact on load power delivery. This exposes power system operation to the possibility of human error and, therefore, increases operating risks.

Though such a conservative operating strategy served the industry well while it was economically possible to do so, it would seem that the electric power industry's prevalent transmission strategy today is wasteful of

Keywords: Electric power transmission; Reliability; Symmetric operation; Asymmetric operation.

expensive transmission equipment, costly in terms of loss of potential revenue, and stressful to both power system equipment and operators. In the light of the considerable pressures on electric utilities due to deregulation, greater environmental awareness, normal load growth, and (as shown later in this article) symmetrical operation expose the system to numerous risks while wasting valuable transmission capacity. This was particularly shown to be true in the August 2003 blackout in the northeastern United States, and Canada in which many lines were switched out due to single-phase faults.[1]

Asymmetric Operation of Power Transmission Systems

Asymmetric operation is essentially defined as the operation of a 3-phase transmission line as three independently operated entities.[2] In this approach, a 3-phase line can be operated with one or two phases out of service for single-line transmission corridors, or with one, two, or even three phases out of service in the case of multiple-line corridors. For this to occur, the strategy implements three operational objectives:

1. Upon entering asymmetric operation, undesirable voltages and currents are "contained" within the affected corridor.
2. From the system perspective, the corridor appears to operate symmetrically at both extremities.
3. From the system perspective, the faulted corridor returns to its precontingency electrical state (i.e., in terms of impedance, voltages, and currents) and maintains precontingency power flows.

Because the post-contingency system is electrically indistinguishable from its precontingency state, this eliminates the need for operator-driven postfault remedial measures. Postfault remedial measures are, therefore, built into the strategy.

To enable asymmetric operation as described above, compensating equipment must be introduced into the corridor just as the faulted phase(s) is(are) switched out by means of circuit breaker action. Such compensating equipment can either be conventional, inexpensive passive devices, such as capacitors and reactors equipped with appropriate switching equipment, or more complex and expensive power-electronic devices such as Flexible AC Transmission System (FACTS) controllers. Both approaches have their strengths and weaknesses, and the final choice will often depend on system-specific constraints imposed by the system planner.

The consequence of asymmetric operation is to add flexibility to transmission system operation, increase corridor reliability, and increase security limits (i.e., the amount of power transferred under normal conditions) and thereby improve transmission economics.

BENEFIT OF ASYMMETRIC OPERATION

Statistics of Transmission Line Failures

Most transmission line faults are single-phase faults (varying from 60 to 97% with increasing voltage level).[3,4] This alone is sufficient motivation for studying asymmetric operation, even if the practice of single pole reclosure is successful up to 50% of the time at voltages up to 765 kV. Because it is so simple to define and to employ, the 3-phase fault—often used concurrently with the subsequent loss of major transmission equipment—has long been the industry norm for estimating system performance under difficult conditions. Even so, one must occasionally be reminded of the fact that 3-phase faults have no more than a 1% probability of occurrence and that their use as a criterion is clearly limited in terms of physical significance. Indeed, given such a statistic, one could argue that a symmetrical 3-phase response to normally occurring events and contingencies is an inappropriate response 99% of the time and, therefore, is far from optimal from an operations perspective.

In the past, this criterion served a useful purpose as an umbrella contingency, covering a large number of contingencies and accounting for lack of knowledge of either a specific operating context or of power system dynamics more generally. Asymmetric operation, however, enables power systems to respond surgically to contingencies while subsuming symmetrical response capability when required.

Reliability Analysis

The most frequently used reliability index in transmission planning is the loss of load expectation (LOLE), which is the expected mean of energy not supplied due to the failure of network components.[5] Here, the LOLE is used for estimating and comparing the risk of operating a single- or multiple-line transmission corridor under symmetric or asymmetric operation.

Fig. 1 shows the logic circuits for the reliability analysis of a 3-phase transmission line under symmetrical or asymmetrical operation. In the symmetrical approach, events leading to the loss of any phase results in the loss of all three phases; in this case, the equivalent logical circuit presents the three phases in series. In the asymmetrical approach, the three phases function independently; thus, the equivalent logical circuit presents the three phases in parallel.

If one considers a corridor of N 3-phase lines transmitting a total power value of T, where the probability of successful transmission of each phase of each 3-phase line is p, the expected mean nontransmitted power according to symmetric operation $LOLE_{sym}$ and the expected mean nontransmitted power according to asymmetric operation $LOLE_{asym}$ are[2]

Elec—Elec

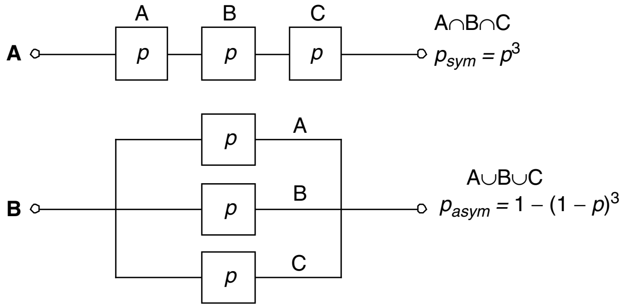


Fig. 1 Equivalent logical circuits of a 3-phase line under (A) symmetric operation and (B) asymmetric operation.

$$\text{LOLE}_{\text{sym}} = f_{\text{ch}} T(1 - p^3)$$
$$\text{LOLE}_{\text{asym}} = f_{\text{ch}} T(1 - p)^3 \qquad (1)$$

where f_{ch} is a load factor that takes average load variations into account. An arbitrary load factor between 50 and 75% is considered to be acceptable in the industry.

The difference between the LOLE of symmetric and asymmetric operation, ΔLOLE, yields the benefit of asymmetric operation, which can be evaluated at the energy generation cost

$$\begin{aligned}\text{LOLE}_{\text{sym}} - \text{LOLE}_{\text{asym}} &= \left[(1 - p^3) - (1 - p)^3\right] f_{\text{ch}} T \\ &= 3p(1 - p) f_{\text{ch}} T \end{aligned} \qquad (2)$$

Eq. 2 shows that ΔLOLE is always greater than zero (ΔLOLE > 0) for $0 < p < 1$. This means that if the probability of successful transmission of the energy from the sending end to the receiving end is the same for all phases, the risk of nontransmission is always higher in the case of symmetric operation.

Example 1 Asymmetric Operation of a 2-Line, 400 kV, 300 km Corridor

The probability of nontransmission can be evaluated as 0.133% per 100 km of line.[6] Thus, for each phase with an equivalent length of 300 km, one has: $q = 1 - p = 0.004$ and $p = 0.996$. The benefit of asymmetric operation is obtained from Eq. 2 as ΔLOLE = $0.01195 f_{ch} T$. Assuming a generation cost of 2500 \$/kW and a load factor of 75%, the benefit of the asymmetric approach over the symmetric approach is approximately 23 \$/kW.

IMPLEMENTATION

Two distinct cases must be addressed for the purpose of implementing asymmetric operation: (1) the multiple-line corridor and (2) the single-line corridor. As previously pointed out, either of two implementation strategies can be employed: (1) conventional devices (i.e., passive LC elements with electromechanical switches or circuit breakers) or (2) power-electronic devices.[7] Though the latter incorporate significant advantages, including rapid response and precise control, the following sections focus on the use of conventional devices due to their lower cost.

Multiple-Line Corridor

A lossless, uncoupled, lumped-parameter transmission line model is used as a starting point for quantifying the compensation strategy; this simplifies the analysis and focuses on the underlying concepts while leading to a reasonable estimate of the capacity and cost of the required compensating equipment. Though more complex distributed parameter line models enable such factors as line resistance to be thoroughly accounted for, the results presented in the following sections are remarkably precise because line resistance is typically very low.[2]

Compensating Impedances

The design of the compensation strategy begins by considering a corridor of N parallel lines ($N \geqq 2$). The corridor, therefore, includes N instances of each phase, where a_i, b_i, and c_i, respectively, refer to the A, B, and C phases of the ith line. Let us consider the case with L of the individual a_i-phases out of service. The problem is to determine the conditions for which the power transmitted on all of the N individual a_i-phases in symmetric mode is equal to that of $N - L$ compensated remaining a_i-phases in asymmetric mode and then deduce the values of the compensating impedances.

Fig. 2A shows the equivalent circuit of the N parallel a_i-phases working in symmetric mode, all individual phases being in service. V_S and V_R are, respectively, the line-to-line rms voltages of the sending and the receiving ends of the transmission system, and X_p and B_p are, respectively, the series impedance and the shunt susceptance of each a_i-phase.

With L of the a_i-phases out of service due to faults, the equivalent circuit of the $N - L$ remaining sound a_i-phases in parallel in asymmetric mode is shown in Fig. 2B. X_S and B_C are, respectively, the series impedance and the shunt susceptance of the compensating devices for each a_i-phase, defined as follows:

$$X_S = -\frac{L}{N} X_p$$
$$B_C = \frac{L}{N - L} B_p \qquad (3)$$

With these compensating elements, the asymmetrically operated compensated corridor A-phase with L of

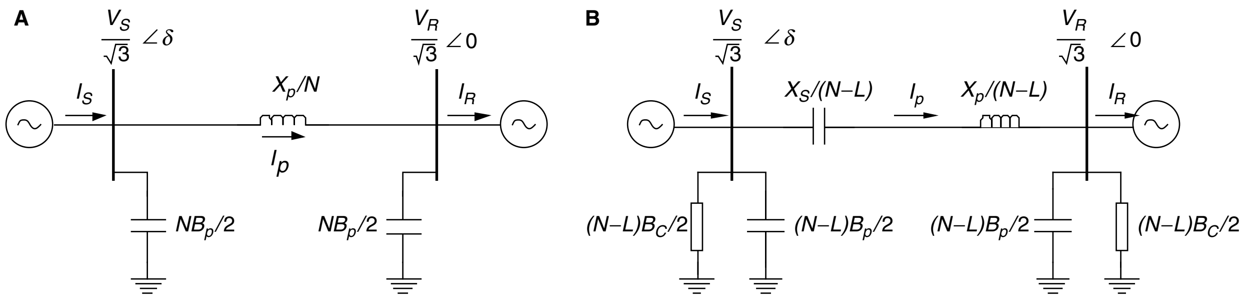


Fig. 2 Equivalent circuit of one phase of a corridor in symmetric and asymmetric operation: (A) corridor A-phase in symmetric operation, consisting of N individual a_i-phases; (B) corridor A-phase in asymmetric operation, consisting of $N-L$ individual a_i-phases.

the a_i-phases out of service has the same electrical characteristics and carries the same power as the original symmetrically operated corridor A-phase with N operational a_i-phases.

Installed Reactive Power

The total installed reactive power for series compensation Q_{Tseries} is calculated assuming that series-connected reactive power is available to every phase of every line. The total installed series reactive power in the $3N$ phases of the corridor is

$$Q_{\mathrm{Tseries}} = 3NX_S\left(\frac{I_p}{N-L}\right)^2 = -3\,\frac{L}{(N-L)^2}\,X_p I_p^2 \qquad (4)$$

where I_p is the rms value of the current in the equivalent line, X_S is the impedance of the series compensating device, and X_p is the series impedance of each a_i-phase.

The value of the total installed shunt reactive power Q_{Tshunt} is again calculated based on the fact that any phase of any line can be lost. There is no need to compensate for any particular a_i-phase, however, as shunt compensation can be installed on the sending-end and receiving-end buses. Consequently, the total reactive power installed for shunt compensation is given by:

$$Q_{\mathrm{Tshunt}} = 3(N-L)B_C V_N^2 = 3LB_p V_N^2 \qquad (5)$$

where B_C is the shunt susceptance of the compensating devices for each a_i-phase; B_p is the shunt susceptance of each a_i-phase; and V_N is the rms phase-to-ground voltage, assumed to be the same at the two ends of the corridor.

Example 2 Two-Line, 400 kV, 300 km Corridor with Lossless Lines (Fig. 3)

In this example, the methodology is applied for the particular case of two 400 kV, 300 km, lossless, transposed lines loaded to their surge impedance loading

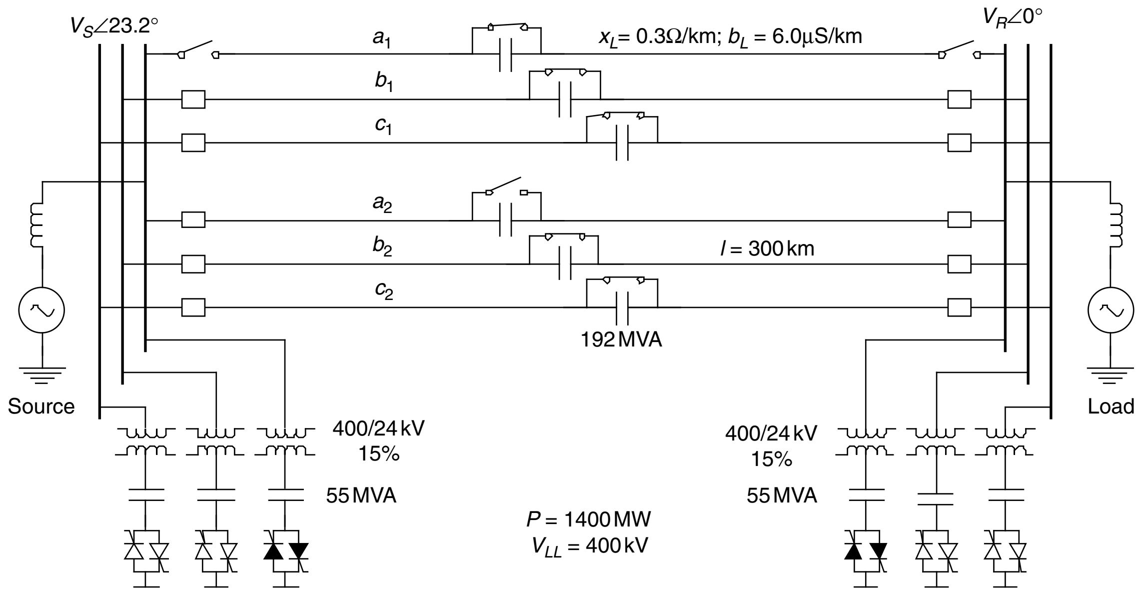


Fig. 3 Compensation scheme for asymmetric operation of a 2-line, 400 kV, 300 km corridor (assuming a lossless conductor model).

(SIL = 1400 MW) and equipped to sustain the loss of three different phases on either of the two 3-phase lines. Fig. 3 shows this corridor with one phase out of service on one line. The unit length parameters are $x_L = 0.3$ Ω/km and $b_L = 6.0$ μS/km. Thus, for 300 km, $X_L = 90$ Ω, and $B_L = 1.80 \times 10^{-3}$ S.

For $L = 1$, Table 1 gives the required reactive power resulting from the application of Eqs. 4 and 5. For series compensation, a total 1150 Mvar is required. The total need for shunt compensation, including 15% reactive impedance representing Static Var System (SVS) transformer losses, is 330 Mvar.

Remarks

The case developed above for N lines with L of the a_i, b_i, or c_i phase out of service is theoretical. Indeed, the case of $L > 1$ has such a low probability of occurrence that it is questionable whether such a contingency need ever be considered. Additionally, as L goes from 1 to 2, compensation requirements and operational complexity are far greater, as one must provide suitable reactive power sources and associated switching for all values of L.

For these reasons, it seems practical to design for the loss of only one a_i-, b_i-, or c_i-phase. Even so, the case of $L = 1$ covers the loss of 1 a_i-phase, 1 b_i-phase, or 1 c_i-phase for any combination of lines, or all three phases of a single line. In the case of a 2-line corridor, the maximum corridor power transfer will be maintained after sustaining up to three single-phase contingencies for any combination of phases on the two different lines. After having sustained a third contingency, however, power transfer would normally be reduced for security considerations. As can be seen, this goes beyond the 3-phase $N - 1$ criterion traditionally used to establish security limits, where N is the number of 3-phase lines in a corridor. Asymmetric operation thus maintains full corridor capacity under challenging circumstances while providing the system operator the precious time required to restore it to its precontingency physical state.

Single-Line Corridor

The solution proposed for the single-line case is based on the application of symmetrical components to balance the resulting 2-phase or 1-phase transmission after the loss of one or two phases, respectively. The currents of the negative and zero sequences resulting from a 1-phase-open or 2-phase-open situation are not negligible and must be filtered or compensated so as to guarantee an adequate asymmetric operation. As the single-phase fault has the highest probability of occurrence, the loss of two phases is not considered here.

Fig. 4 shows the principle of asymmetric operation of a 3-phase line with phase a out of service. In normal symmetric operation, the two connected networks and the line are perfectly balanced. During asymmetric operation with one phase open, three basic compensating elements must be introduced to rebalance the voltages and currents:

1. Series compensation of the sound phases (phases b and c in Fig. 4) to lower the series reactance, maintain the same angular spread, and thus ensure the flow of the precontingency power transfer. Series compensation can be supplied by means of either conventional capacitors or series-controlled voltage sources employing power-electronic devices. The compensating elements can be placed at the sending end, the receiving end, or the center of the line.
2. Zero-sequence filters, at each end of the line, to afford a low-impedance path for the zero sequence current. Many designs can be used to implement such filters:
   - a single T, zigzag, or Δ-Y transformer
   - combinations of passive LC elements that may be variable to reflect large load variations
   - controlled shunt current sources made up of power-electronic devices.
3. Negative sequence compensators at each end of the line to eliminate the negative-sequence current by injecting an opposite current of the same magnitude. Here again, the negative sequence compensators can be constructed of passive LC elements connected in delta or star, or of controlled shunt current sources employing power electronic converters. Depending on the planning criteria, the negative-sequence compensator and zero-sequence filter at each end of the line can be grouped together.[2]

Table 1 Total reactive power requirements for the asymmetric operation of a 2-line, 400 kV, 300 km corridor equipped to sustain the loss of any three phases in the corridor (assuming a lossless conductor model)

Total transmitted power $P = 1400$ MW	
Reactive power Q (Mvar)	
Series compensation	1150 Mvar
Shunt compensation	330 Mvar

Example 3 Asymmetric Operation of a Single-Line, 120 kV, 100 km Corridor (Fig. 5)

A 3-phase, 120 kV, 100 km line with parameters $r = 0.061$ Ω/km, $x_L = 0.3644$ Ω/km, $b_L = 4.54 \times 10^{-6}$ S/km is considered. The load power is 50 MW. Fig. 5 and Table 2 show the calculation results and the reactive power needed for asymmetric operation of the line with a phase open. Compensation of the positive sequence requires three

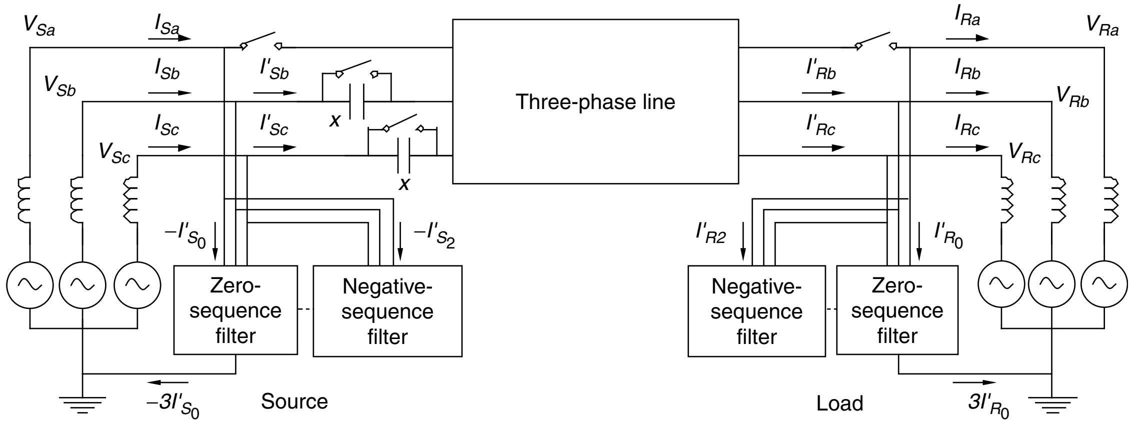


Fig. 4 Concept of asymmetric operation of a single-line corridor.

capacitors of 2 Mvar each for a total 6 Mvar. The compensation of the negative sequence requires 68 Mvar for the two compensators at the ends of the line. Filtering the zero sequence requires 16 Mvar.

Remarks

The values of the passive LC elements of the negative-sequence compensator depend on the actual load, and it may be necessary to adjust them for large variations of the load from SIL (design load). Adjusting the grounding transformers of the zero-sequence filters according to the actual load is not necessary because impedances are designed for maximum load conditions.

In a sense, this solution is an extension of the principle of load compensation.[8] As pointed out earlier, FACTS controller-based implementations can provide better and more rapid control but cost more than conventional solutions.

FINANCIAL ANALYSIS

The LOLE can also be expressed as the loss of revenue related to the energy not delivered in 1 year; such a

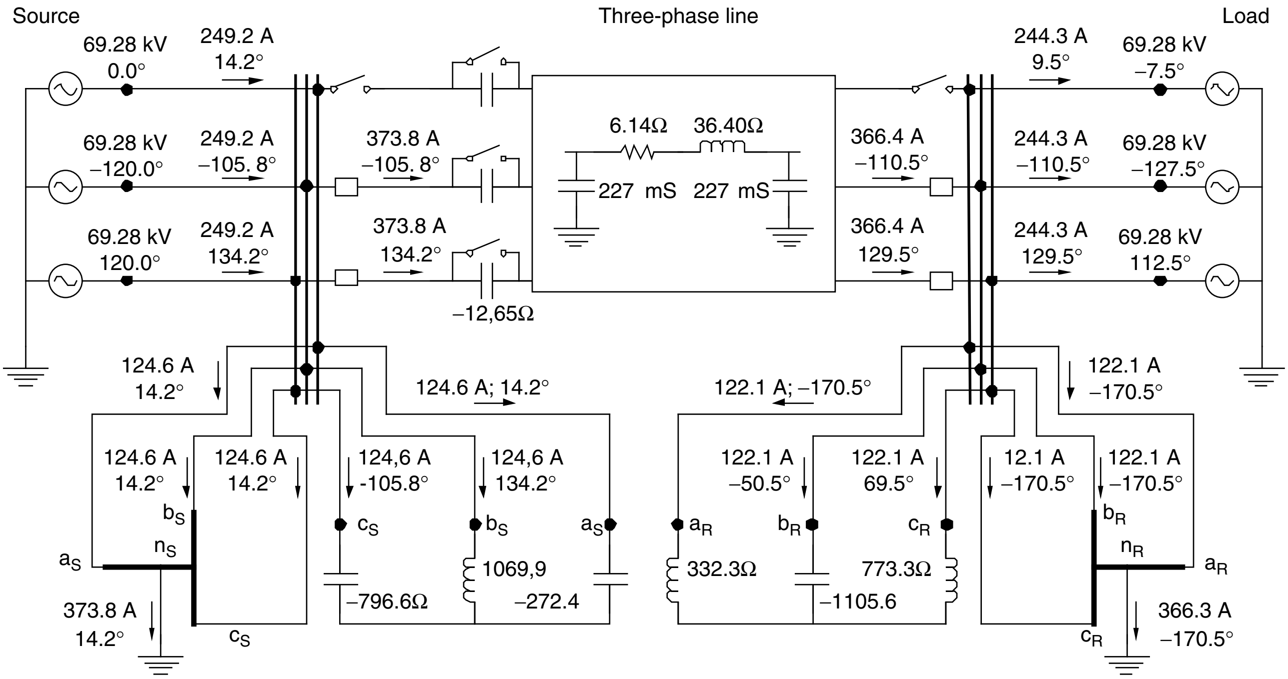


Fig. 5 Asymmetric operation of a single-line, 120 kV, 100 km corridor with one phase open (assuming the use of conventional compensating devices).

Table 2 Calculated results of the operation of a single-line, 120 kV, 100 km corridor with one phase open, using conventional compensating devices

		Sending end	Receiving end 50 MW 16.4 Mvar
Characteristics of the line: $R_L = 6.14\ \Omega$; $X_L = 36.4\ \Omega$; $B_c = 4.54 \times 10^{-4}$ S; $X_{c=} 12.65\ \Omega$ (3×2 Mvar)			
System voltages	E_{ab} (kV)	120.0∠30.0°	120.0∠22.6°
Line currents before asymmetric operation	I_a (kA)	0.249∠9.3°	0.244∠2.1°
Line currents during asymmetric operation	I'_b (kA)	0.374∠−106°	0.366∠−110.5°
	I'_c (kA)	0.374∠134°	0.366∠129.5°
Negative-sequence compensator			
Compensating impedances	X_a (Ω)	−272.4	332.3
	X_b (Ω)	1069.9	−1105.6
	X_c (Ω)	−797.6	773.3
Reactive power of the compensator	S_a (Mvar)	−4.2	5.2
	S_b (Mvar)	16.6	−17.2
	S_c (Mvar)	−12.4	12.0
Total	S_T(Mvar)	33.2	34.4
Zero sequence filter			
Transformer capacity	S_{Tg}(MVA)	8.1	8.1

figure can then be used to determine how long it takes to pay for the equipment required to implement either asymmetric operation or some other alternative, such as adding another line. The purpose of this section is to show that asymmetric operation is a practical and economical alternative for improving transmission system capacity.

Retrofit of Existing Corridors

The benefit of asymmetric operation is obtained by comparing (1) the total investment cost of retrofitting an existing 3-phase corridor for asymmetric operation and (2) the LOLE of 3-phase symmetrical operation—expressed in dollars—which is offset by means of asymmetric operation. Initially, the annual savings generated by asymmetric operation can be used to reimburse the equipment investment; after this investment has been repaid, this represents additional revenue to the utility. Table 3 summarizes the financial analysis for three 400 kV, 150 km asymmetrically operated corridors involving, respectively, one, two, and three transmission lines. For purposes of comparison, one also finds both the cost of a new line on this table and the cost of implementing a typical FACTS-based compensation strategy.[2]

For the single-line case, the cost of the LOLE is lower than the investment cost of asymmetric operation (and much lower than the cost of building a new line). The payback period, therefore, is greater than 1 year. The cost of LOLE for this case, however, does not take into account the larger social and economic costs associated with the total loss of the power supply in the symmetric operation mode, which are not addressed here.

For two or three lines, the investment cost of asymmetric operation using conventional devices is lower than the cost of the LOLE: in this case, the payback time of asymmetric operation is less than 1 year.

In a general way, one sees that the cost of implementing asymmetric operation by retrofitting multiple lines is always lower than in the single-line case because one exploits the transmission equipment already in place. In other words, part of the equipment required to implement asymmetric operation is already there!

As a final note, all three cases provide remarkable performance in relation to the standard 3-phase $N-1$ security criterion.[9] Case 1, when operated asymmetrically, respects the single-phase $N-1$ criterion, whereas symmetrical operation of this same line is incapable of respecting any $N-1$ criterion, either 1-phase or 3-phase. Cases 2 and 3, when operated asymmetrically, are capable of respecting the full 3-phase $N-1$ criterion in addition to a single-phase $N-1$ criterion. All this translates into high transmission reliability and improved transmission economics.

Table 3 Comparing the cost of three options at 400 kV for reducing the risk of nontransmitted energy (costs are expressed in millions of dollars Canadian)

	Case 1	Case 2	Case 3
Number of lines	1	2	3
Voltage (kV)	400	400	400
Line length (km)	150	150	150
Transmitted power (MW)	700	1400	2000
Probability of failure of one phase[6]	0.002	0.002	0.002
Load factor (%)	75	75	75
Cost of power not supplied: LOLE (M$CA)	8	16	23
Asymmetric operation with conventional devices			
Investment Cost (M$CA)	48	15.0	12.1
Payback time (years)	6.0	0.93	0.53
Asymmetric operation with FACTS devices			
Investment cost (M$CA)	70	42.8	34.7
Payback time (years)	8.8	2.7	1.5
New line			
Investment cost (M$CA)	60	60	60
Payback time (years)	7.5	3.8	2.6

New Corridors

Three voltage scenarios (345, 500, and 735 kV; see Table 4) have been selected to compare the performance of different transmission options under symmetric and asymmetric operation for planning new transmission capacity from the point of view of the $N-1$ criterion. Each scenario compares the cost of a 2-line, symmetrically operated transmission corridor to a 1-line, asymmetrically operated corridor of the same capacity, both of which respect an $N-1$ criterion. In each case, the $N-1$ criterion is interpreted as follows: For a symmetrically operated system, $N-1$ represents the loss of a 3-phase line; for an asymmetrically operated system, $N-1$ represents the loss of a single phase.

In each voltage scenario, both cases have the same capacity, as normal planning and operating criteria generally load a double-circuit corridor to no more than the SIL of a single line for reliability purposes. This ensures that the corridor respects the 3-phase $N-1$ criterion—in this case, the loss of one of the two lines without loss of load.

Because the operational reliability of these asymmetric and symmetric scenarios is essentially identical (as they respect their respective $N-1$ criterion), LOLE is not an appropriate basis for comparison. To compare the three scenarios, one must consider their respective investment costs.

Cost Analysis

Table 5 presents a summary of the costs associated with the construction of a 300 km, symmetrically operated 2-line corridor at 345 kV, 500 kV, and 735 kV; it also presents a summary of those associated with the construction of a 300 km, asymmetrically operated single-line corridor at each of these respective voltage levels. These estimates are based on the use of conventional elements and include rights of way.

In all three scenarios, the cost of a single, asymmetrically operated 300 km line is less than that of two symmetrically operated 300 km lines. This is because the cost of the reactive power for implementing asymmetric operation of a 300 km line is less than the cost of the additional line required in symmetric operation.

Elec–Elec

Table 4 Transmission system scenarios for comparing symmetric and asymmetric operation

	Scenario 1	Scenario 2	Scenario 3
Transmitted power	450 MW	1000 MW	2200 MW
Symmetric operation	345 kV; 2×300 km lines	500 kV; 2×300 km lines	735 kV; 2×300 km lines
Asymmetric operation	345 kV; 1×300 km line	500 kV; 1×300 km line	735 kV; 1×300 km line

Table 5 Costs associated with symmetric and asymmetric operation for the three scenarios of Table 4 (costs are expressed in millions of dollars Canadian)

	Scenario 1	Scenario 2	Scenario 3
Transmitted power (MW)	450	1000	2200
Symmetric operation			
Voltage level; Nb. lines	345 kV; 2 lines	500 kV; 2 lines	735 kV; 2 lines
Total cost (lines) (M$CA)	120	240	360
Asymmetric operation			
Voltage level; Nb. Lines	345 kV; 1 line	500 kV; 1 line	735 kV; 1 line
Cost of the lines (M$CA)[10]	60	120	180
Cost of compensation (M$CA)	26.8	59.2	128.6
Total cost (M$CA)	86.8	179.2	306.6

Effect of Line Length

In the above example, line length was set somewhat arbitrarily at 300 km even though all scenarios are technically realistic. There is considerable merit, however, in comparing the costs of a 1-line, asymmetrically operated corridor as a function of distance with respect to those of a 2-line symmetrically-operated corridor. This was examined, therefore, in the case of the three voltage scenarios considered above and for line lengths ranging from 1 to 400 km.

As Fig. 6 shows, the costs of both symmetric and asymmetric operation vary linearly with line length but at different rates. For short lines, the cost of asymmetric operation is higher because the cost of shunt compensation is predominant. For long lines, the cost of symmetric operation is higher because the cost of the lines increases more quickly than the cost of compensation. As shown, there exists a point for each scenario at which, between 100 and 200 km, an asymmetrically operated 1-line corridor costs less to build than a symmetrically operated 2-line corridor of the same voltage. Such cost behavior is similar to that found in comparisons of AC and DC transmission corridors, where, beyond a certain point, DC transmission is less costly than the equivalent AC solution.

CONCLUSION

Asymmetric operation transforms power system planning and operation by virtue of the greater flexibility available in finding solutions to specific challenges. Though actual implementations will generally require detailed simulation and engineering of components and systems for such purposes as insulation coordination, protection, control, security, and reliability, the numerous examples presented here show that the concepts are applicable to any voltage level.

In the case of multiple-line corridors, it has been shown that conversion to asymmetric operation increases the corridor availability, reduces the loss of load expectation, and increases the secure power transfer limit (as each and every line can be operated

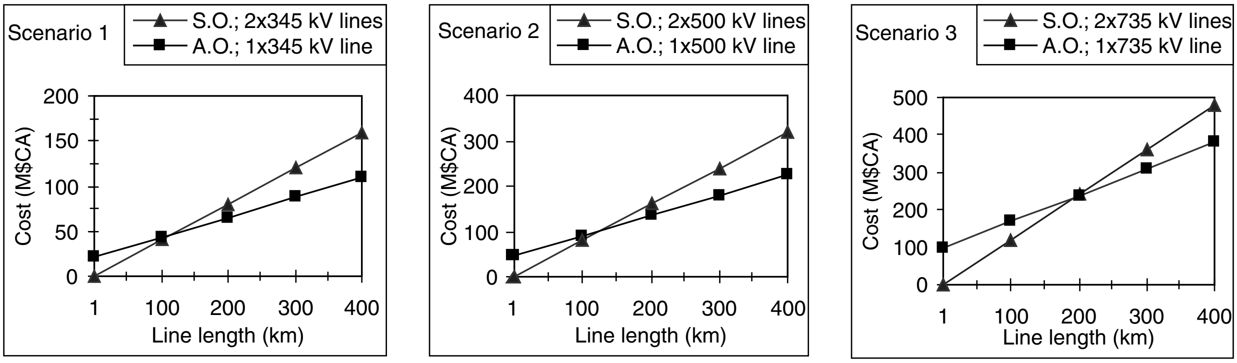


Fig. 6 Cost comparison of symmetric (S.O.) and asymmetric (A.O.) operation for three scenarios as a function of line length (costs are expressed in millions of dollars Canadian).

at its SIL) while respecting existing 3-phase $N-1$ security criteria if need be. In the case of new transmission corridors, point-to-point transmission, interarea network interconnections, or ring-type transmission grids feeding large metropolitan areas, asymmetric operation has such a positive impact on the reliability of single-line transmission systems that it redefines one's outlook on such fundamental issues as choice of voltage, number of lines, the contingency that defines one's security criterion, the amount of power transferred securely, and the amount of land required for transmission rights of way. In short, asymmetric operation redefines the reliability, environmental impact, and economics of the planning and operation of AC power transmission.

REFERENCES

1. U.S.-Canada Power System Outage Task Force. Final Report on the August 14, Blackout in the United States and Canada: Causes and Recommendations. April 2004. Available at: www.nrcan-rncan.gc.ca/media/docs/final/finalrep_e.htm (accessed on October 2005).
2. Marceau, R.J.; Sana, A-R.; McGillis, D.T. *Asymmetric Operation of AC Power Transmission Systems*; Presses Internationales Polytechnique: Montreal, Quebec, Canada, 2006.
3. Blyakov, N.N.; et al. Utilisation du Réenclenchement Monophasé Automatique dans un Grand Réseau EHT Comportant des Lignes 1200 kV. CIGRE **1990**, 34–207.
4. Automatic reclosing of transmission lines. In *IEEE Transactions on Power Apparatus and Systems*, IEEE Power System Relaying Committee Report, 1984; PAS-103(2).
5. Anders, G.J. *Probability Concepts in Electric Power Systems*; Wiley: New York, 1990.
6. Canadian Electrical Association. Forced Outage Performance of Transmission Equipment for the Period of January 1, 1994 to December 31, 1998. February 2000.
7. Hingorani, N.G.; Gyugyi, L. Understanding FACTS. IEEE Press: New York, Power Systems Engineering, 2000.
8. Gyugyi, L.; Otto, R.A.; Putman, T.H. Principles and applications of static, thyristor-controlled shunt compensators. IEEE Trans. Power Apparatus Systems **1978**, *PAS-97* (5), 1935–1945.
9. Marceau R.J. et al. Power system security assessment: A position paper. CIGRE Task Force 38.3.12 Final Report. June 30, 1997.
10. Picard, B.; Galiana F.D.; McGillis, D. A knowledge-based system for the structural design of high-voltage lines. Proceedings of IEEE-CCECE'1999, Edmonton, Alberta, Canada, 1999.

Electric Supply System: Generation

Jill S. Tietjen
Technically Speaking, Inc., Greenwood Village, Colorado, U.S.A.

Abstract
The electric utility system is comprised of three major components: the generation system, the transmission system, and the distribution system. The generation system, where the electricity is produced, is comprised of power plants, also called generating units. Generation in the United States is produced at facilities categorized as conventional and renewable resources. Conventional resources are those in which the fuel is burned, and include coal, nuclear, natural gas, oil, and diesel power plants. Renewable resources are those in which the "fuel" consumed to produce electricity is replenished by nature or can be replenished by humankind. Generation resources in the United States are predominantly conventional resources, although renewable resources are in operation and providing more and more electricity every year.

INTRODUCTION

The electric utility system is comprised of three major components: the generation system, the transmission system, and the distribution system. The generation system, where the electricity is produced, is comprised of power plants, also called generating units. The transmission and distribution systems both consist of wires and other equipment that carry the electricity from the power plants (or generation sources) to the homes and businesses where we consume the power. The transmission lines and towers move large amounts of power from the power plants to large population areas (cities and towns), where it is converted at substations to lower voltages. The distribution system carries the power at lower voltages from those substations to our actual houses and businesses.

Generation in the United States is produced at facilities categorized as conventional and renewable resources. Conventional resources are those in which the fuel is typically burned, and include coal, nuclear, natural gas, oil, and diesel power plants. Renewable resources are those in which the "fuel" consumed to produce electricity is replenished naturally by nature or can be replenished by humankind; these resources include hydroelectric, geothermal, solar, wind, and biomass. Generation resources in the United States are predominantly conventional resources, although renewable resources are in operation and are providing more and more electricity every year in the United States.

Keywords: Power plants; Conventional resources; Renewable resources; Coal; Nuclear; Natural gas; Oil; Diesel; Wind; Solar; Geothermal; Biomass; Turbine; Generator.

CONVENTIONAL RESOURCES

The fuels used in conventional resources to generate electricity in the United States include coal, nuclear, natural gas, oil, and diesel. Coal, natural gas, and oil are burned in a boiler that heats water to produce steam. Nuclear power plants generate steam from the heat given off by nuclear fission. Natural gas is burned to generate electricity in combustion turbines and combined cycle power plants. Oil can also be burned to produce electricity in combustion turbines. Diesel generators tend to have an internal combustion engine that directly turns a generator to produce electricity.

A typical coal-fired power plant works as demonstrated in Fig. 1. The coal is fed into the boiler, where it is burned to heat water to convert the water to steam. At that point in the cycle, the steam is high temperature and high pressure. This steam moves into one or more turbines mounted on the same or on separate shafts; the steam turns the turbine blades. The turbine blades are connected to the turbine shaft, which is connected directly or indirectly to the generator rotor (shaft). As the generator shaft rotates, it produces electricity (see Fig. 2). The steam that exits the turbine is now low pressure and lower temperature, because its pressure and temperature have been reduced in the process of its working to turn the turbine blades. The steam is sent to a cooling tower or condenser, where additional heat is removed so that the steam can be converted back to water. Water cooled through the cooling tower or condenser then can be discharged back into the body of water from which it came—usually, a river or a lake. Or as shown in Fig. 1, water can be pumped directly back to the boiler from the cooling tower or condenser.

Coal-fired generating units currently produce about half of the energy in the United States.[1] Many smaller, older units are still in operation. Newer units tend to be in the size range of 250 MW to more than 1,000 MW.

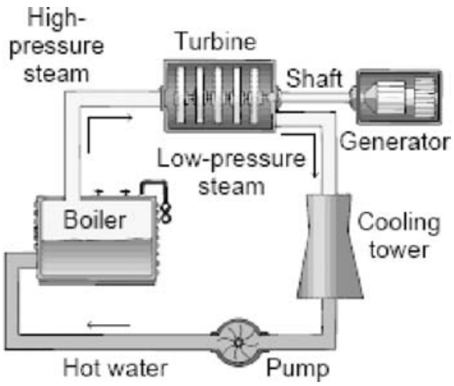


Fig. 1 Schematic diagram of a conventional fossil fueled generating station.
Source: From Wiley (see Ref. 2).

Coal is mined and shipped to the power plants from a broad range of states. The vast majority of coal is mined in Wyoming, West Virginia, Kentucky, Pennsylvania, and Texas.

Coal-fired power plants are very reliable and operate in a manner that is called dispatchable. This means that the amount of power that the facility provides minute to minute is controllable by a human operator and can be increased or decreased at any time, depending on the amount of electricity required by a utility's customers. Typically, the output is controlled by an integrated control system based on automatic inputs and feedback. New coal-fired power plants have a significant amount of pollution control equipment installed that can include some or all of the following: electrostatic precipitators (to control particulate matter), flue-gas desulfurization (to reduce sulfur dioxide emissions), and selective catalytic reduction (SCR) (to reduce nitrogen oxide (NO_x) emissions). Coal-fired power plants also produce carbon dioxide (CO_2) in the combustion process.

Nuclear power plants are in some ways similar to coal-fired power plants and in some ways different. Similarities include that the nuclear fuel heats water to turn it into steam and that turbines turn generators to produce electricity.

In boiling water reactors, there is one water system, as with coal-fired power plants. For pressurized water reactors, however, there are two water circulation systems. As shown in Fig. 3, these are called the primary water loop and the secondary water loop. The water in the primary water loop stays within the containment building, where it is heated by nuclear fission in the reactor and in turn heats the water in the secondary water loop, causing it to turn into steam. The secondary water loop is very similar to the water system in a coal-fired power plant, as it goes through the turbine, is cooled, and then is pumped back through the cycle to be reheated.

Nuclear power plants produce no greenhouse gases in the process of combustion. The fuel and other portions of the power plant are radioactive, however, and require special handling and storage. Nuclear power plants currently account for about 10% of the installed generating capacity in the United States.[1] Because of their high reliability and low fuel cost, nuclear units generally run at full capacity when they are online, and utility dispatchers rarely lower or raise the amount of power being generated from them on a minute-to-minute basis. No new nuclear units have been built in the United States for many years, but as of 2006, several consortia are planning new units that could be constructed and providing electricity by 2015 or later.

Combustion turbine technology burns natural gas (or some other type of liquid fuel) in an internal combustion chamber to heat compressed air.[4] The heated air turns the turbine blades, which turn the generator to produce electricity (see Fig. 4). Combustion turbines are primarily used by utilities to provide energy during the peak period (when the load is at its highest—generally, on hot summer

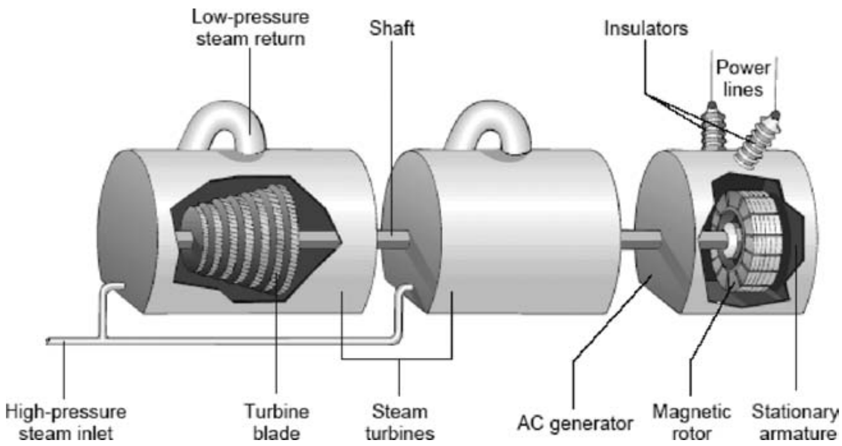


Fig. 2 Turbine-generator configuration.
Source: From Wiley (see Ref. 2).

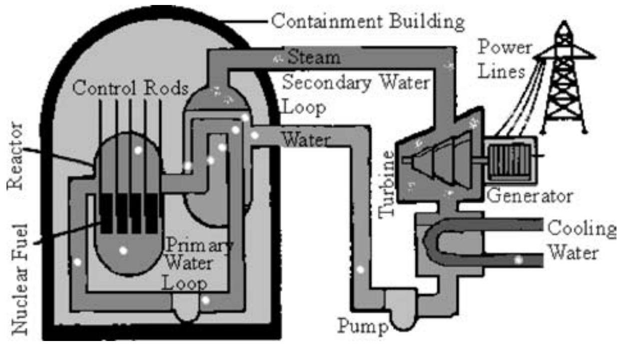


Fig. 3 Diagram of a nuclear power plant: pressurized water reactor.
Source: From Nuclear Reactors (see Ref. 3).

afternoons and cold winter mornings) or during emergencies (such as the loss of a major generating unit or transmission line).

Combustion turbines range in size from about 1 MW to more than 300 MW. New units use selective catalytic reduction in addition to water or steam injection to control NO_x emitted during the combustion process. Combustion turbines also emit CO_2.

In special combined cycle units, much of the heat that is exhausted (not used) in the process of generation from a combustion turbine is captured using a heat recovery steam generator (HRSG). This HRSG uses the waste heat to turn water into steam and produce additional electricity. This combination of combustion turbines and an HRSG is much more efficient than the standard combustion turbine power plant in a stand-alone mode.

The combined cycle unit can be used throughout the day by the electric utility dispatcher, and the amount of generation that it produces can be dispatched (raised and lowered as the electric utility customers' load increases and decreases). Because combined cycle units include a combination of combustion turbines and HRSGs, the total installed capacity for this type of unit ranges from 10 MW to more than 950 MW. Emissions from these units include NO_x and CO_2, and SCR is generally used. Installed capacity in the United States fueled by natural gas—including combustion turbines, combined cycle units, and steam units—comprises about one-quarter of all capacity.[1]

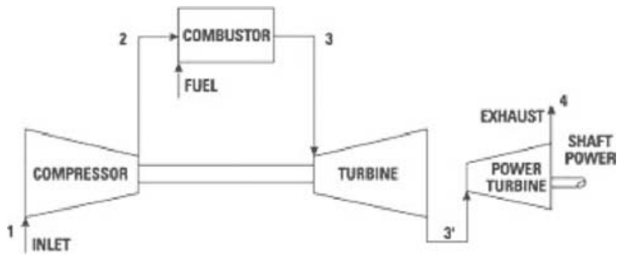


Fig. 4 Combustion turbine schematic.
Source: From Energy Solutions Center (see Ref. 4).

Diesel generators are a reciprocating engine technology that burns diesel fuel. In these engines, an air-and-fuel mixture is burned. These smaller generators are often used as the source of electricity in remote villages and, thus, run 24 h per day. In addition, they are used as emergency backup, particularly at nuclear power plants, and run only for very limited hours in a year, if at all. Diesel generators emit particulates, sulfur dioxide, NO_x, carbon monoxide and CO_2. A small percentage of total U.S. installed generating capacity consists of diesel generators.

RENEWABLE RESOURCES

Renewable resources in the United States have received added attention in the past few years as a means of producing power without the environmental effects associated with the burning of fossil fuels (coal, oil, natural gas, and diesel). Renewable resources in the United States currently used for electricity generation include hydroelectric (conventional and pumped storage), geothermal, wind, solar, and biomass.

Hydroelectric energy (energy that comes from water) has been in use in the electric utility industry since that industry's infancy. Conventional hydroelectric resources use flowing water to move turbines that turn generators, which produce electricity. These facilities can either be run-of-the-river or storage hydro facilities. Run-of-the-river means that as the water flows in the river, it passes through the dam and generates electricity. This type of capacity is not within the control of the utility dispatcher. Storage hydro means that water can be stored behind a dam for some period and then released to produce electricity at the command of the electric utility dispatcher. The dispatcher normally would want to use the water to produce electricity during system peaks. About 8% of the current installed electric generating capacity in the United States is provided by hydroelectric power plants.[1]

Pumped storage hydro takes advantage of significant height differences between an upper reservoir and a lower reservoir. During the day and at the control of the operating utility, water flows downhill and produces electricity. At night, water is pumped back uphill so that it can be used the next day or otherwise in the future. Pumped storage hydro actually requires energy to perform the pumping, and because it is 70%–85% efficient, it actually consumes more energy than it produces. The cost differential between the energy used to pump the water uphill at night and the energy that would be generated by other fuel sources during peak hours that instead is replaced with the water from the pumped storage facility allows this form of electric generating capacity to be cost effective. Pumped storage capacity amounts to just over 2% of installed generating capacity in the United States.[1]

Neither conventional hydroelectric generation nor pumped storage hydro generation emits any greenhouse gases. Environmental concerns related to hydroelectric

Elec–Elec

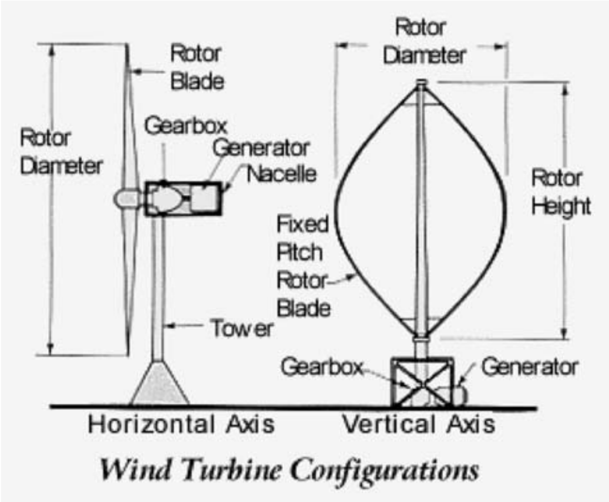


Fig. 5 Wind turbine configurations.
Source: From AWEA (see Ref. 7).

facilities include changes to stream flows, oxygen content of the water, and the disruption of wildlife habitat.[5]

The word geothermal comes from Latin words that mean "heat from the earth." Geothermal resources range from shallow ground to hot water and rock several miles below the Earth's surface, to even farther down, to molten rock known as magma. In the United States, most geothermal resources are located in the Western states, Alaska, and Hawaii.

The three types of geothermal power plants operating today are dry steam plants, flash steam plants, and binary-cycle plants. Dry steam plants directly use geothermal steam to turn turbines. Flash steam plants pull deep, high-pressure hot water into lower-pressure tanks and use the resulting flashed steam to drive turbines. Binary-cycle plants pass moderately hot geothermal water by a secondary fluid with a much lower boiling point than water. This process causes the secondary fluid to flash to vapor, and this vapor then drives the turbines. Geothermal energy emits little to no greenhouse gases and is very reliable (with an average system availability of 95%). Geothermal heat pumps use the nearly constant temperature of the upper 10 ft of the Earth's surface to both heat and cool residences.[6]

Wind energy transforms the kinetic energy of the wind into electrical energy. Wind turbines come in 2 types: vertical-axis (eggbeater style), and horizontal-axis (propeller-style) machines (see Fig. 5). The turbine subsystems include a rotor or blades to convert the wind energy to rotational shaft energy, a nacelle (an enclosure) to cover the drive train (which usually includes a gearbox and a generator), a tower to support the rotor and drive train, and electronic equipment to connect the facility to the utility's transmission grid.

Wind energy produces no greenhouse gases but does raise concerns with respect to noise and presenting a danger to migratory birds and bats. Another concern about wind energy generation is that power is produced only when the wind blows and, thus, is not dispatchable. The wind may not be blowing at the times when it is most needed (i.e., on hot summer afternoons). Unlike fossil fuel-fired power plants, the maximum usage expected from wind-energy turbines would be that they generate energy at about 30%–40% of the maximum amount of energy that would be possible if the wind were blowing every hour.[7]

Installed wind energy capacity reached 9149 MW at the end of 2005 (out of total installed U.S. generating capacity of more than 960,000 MW). The states with the most wind turbines installed, in decreasing order, are California, Texas, Iowa, Minnesota, Oklahoma, New Mexico, Washington, Oregon, Wyoming, and Kansas. Much new capacity is expected to come online during 2006 and 2007 due to the availability of the Federal Production Tax Credit.[1,8]

The rays from the sun can be harnessed as what is called solar power. A variety of forms are possible for utility application in the form of power plants. Concentrating solar power (CSP) technologies include dish/engine systems, trough systems, and power towers. CSP systems use reflective materials such as mirrors to concentrate the sun's energy. The concentrated heat energy then is converted to electricity. Photovoltaic (PV) systems (so-called solar cells that are familiar from America's space program) are built into arrays that directly convert sunlight into electricity. To date, solar facilities that generate electricity are primarily located in California. Much research and development is ongoing to make solar energy more cost effective for a wide range of applications.

Solar generation facilities do not emit any greenhouse gases, but current facilities use up many more acres of land compared with a conventional resource for the same amount of electricity output. Solar generation typically produces electricity only when the sun is shining unless some form of energy storage device has been installed.[9]

Biomass electric generation is the largest source of renewable energy that is not hydroelectric. Biomass means any plant-derived organic matter available on a renewable basis, including dedicated-energy crops and trees, agricultural food and feed crops, agricultural crop wastes and residues, wood wastes and residues, aquatic plants, animal wastes, and municipal wastes. Waste energy consumption generally falls into categories that include municipal solid waste, landfill gas, other biomass, and other. Other biomass includes agriculture byproducts and crops; sludge waste; tires; and other biomass solids, liquids, and gases. Biofuels being developed from biomass resources include ethanol, methanol, biodiesel, Fischer–Tropsch diesel, and gaseous fuels such as hydrogen and methane.[10]

Biomass power systems fall into 4 categories: direct-fired, co-fired, gasification, and modular systems. Most biomass systems today are direct-fired systems that are quite similar to most fossil fuel-fired power plants where

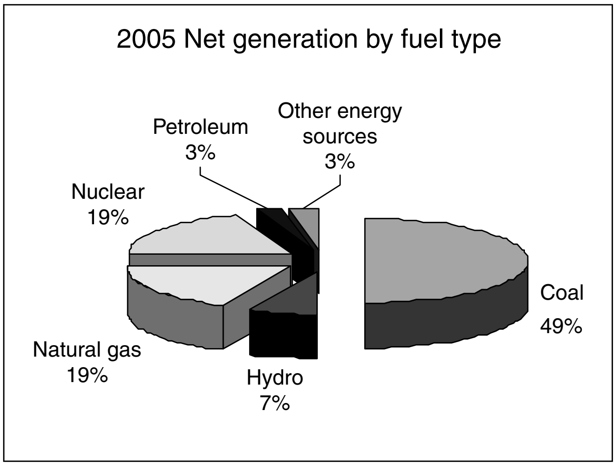


Fig. 6 2005 net generation by fuel type.
Source: From Energy Information Administration, Electric Power Monthly, March 2006: With Data for December for 2005, DOE/EIA-0226 (2006/03), www.eia.doe.gov/cneaf/electricity/epm/epm_sum.html, accessed April 2006.

the biomass is burned, providing heat to turn water into steam, which turns blades in a turbine that turns a generator and produces electricity. The direct-fired systems produce many of the same products of combustion as do coal-fired generating units and require much of the same equipment for cleaning up the byproducts. Direct-fired systems tend to be dispatchable, although some are constrained by the amount of fuel available to the facility.

Co-firing involves substituting biomass for a portion of coal in an existing power plant boiler. Co-firing is a much less expensive way to burn biomass than building a new biomass power plant. Biomass gasifiers heat the biomass in equipment in which the solid biomass breaks down to a flammable gas. The gas can be cleaned and filtered to remove problem elements and then burned much like natural gas in a combined cycle unit. Modular systems are either direct-fired or use gasifiers in small-scale situations, such as for villages, farms, and small industry.[11]

Elec–Elec

GENERATION IN THE UNITED STATES

Coal is the predominant fuel used for electricity generation in the United States. Fig. 6 shows the actual generation for 2005, with coal producing 49.9% of the electricity consumed during the year. The large majority of the remainder of electricity produced in 2005 came from nuclear and natural gas. Hydroelectric is large enough to be identified separately, but the remaining renewable resources are included in the other energy sources.[12]

CONCLUSION

The generation system is one of the three components of an electric utility's system. That generation (the power plants) is comprised of both conventional and renewable resources. Conventional generation resources including coal, nuclear, natural gas, and oil continue to provide the large majority of electricity produced in the United States. Renewable resources are growing but represent a small slice of total electricity production.

REFERENCES

1. Energy Information Administration. Electric Power Annual (with date for 2004), released November 2005. Available at: www.eia.doe.gov/cneaf/electricity/epa/epa_sum.html (accessed April 2006).
2. Bloomfield, L.A. *How Things Work: The Physics of Everyday Life*; Wiley: New York. Available at: www.howthingswork.virginia.edu/supplements/electric_power_-generation.pdf (accessed April 2006).
3. Nuclear Reactors. University of Wisconsin Nuclear Reactor Tour, College of Engineering, University of Wisconsin, Madison. Available at: http://reactor.engr.wisc.edu/power.html (accessed April 2006).
4. Energy Solutions Center (ESC). Combustion Turbine. Available at: www.energysolutionscenter.org/DistGen/Tutorial/CombTurbine.htm (accessed April 2006).
5. California Energy Commission. Hydroelectric Power in California. Available at: www.energy.ca.gov/electricity/hydro.html (accessed April 2006).
6. U.S. Department of Energy. Geothermal Basics Overview. Energy Efficiency and Renewable Energy Geothermal Technologies Program. Available at: www1.eere.energy.gov/geothermal/printable_versions/overview.html (accessed April 2006).
7. AWEA. Wind Energy Basics. Available at: www.awea.org/faq/tutorial/wwt_basics.html (accessed April 2006).
8. AWEA. Wind Energy Projects Throughout the United States of America. Available at: www.awea.org/projects (accessed April 2006).
9. U.S. Department of Energy. Concentrating Solar Power. Energy Efficiency and Renewable Energy, Solar Energy Technologies Program. Available at: www1.eere.energy.gov/solar/printable_versions/csp.html (accessed April 2006). U.S. Department of Energy. Solar Energy Topics. Energy Efficiency and Renewable Energy. Available at: www.eere.energy.gov/RE/solar.html (accessed April 2006).
10. U.S. Department of Energy. Biomass Topics. Energy Efficiency and Renewable Energy. Available at: www.eerer.energy.gov/RE/biomass.html (accessed April 2006). Energy Information Administration. Landfill Gas. Available at: www.eia.doe.gov/cneaf/solar.renewables/page/landfillgas/landfillgas.html (accessed April 2006).
11. U.S. Department of Energy—Energy Efficiency and Renewable Energy Biomass Program. Electrical Power Generation. Available at: http://eereweb.ee.doe.gov/biomass/printable_versions/electrical_power.html (accessed April 2006).
12. Energy Information Administration. Electric Power Monthly, March 2006: With Data for December for 2005. DOE/EIA-0226 (2006/03). Available at: www.eia.doe.gov/cneaf/electricity/epm/epm_sum.html (accessed April 2006).

Electricity Deregulation for Customers

Norm Campbell
Energy Systems Group, Newburgh, Indiana, U.S.A.

Abstract
Deregulation of the United States electric industry has been a roller coaster ride with roots in the 1920's and continuing as a work in progress to present day. From the birth of retail electricity markets and the initial regulations provided through PUCA in 1935, the electric industry has fueled the growth of the United States. Since PURPA in the 1970's, the Energy Policy Act of 1992 and most recent Energy Policy Act of 2005, regulators have been attempting to balance the yoke of regulation with that of open competition while continuing the history of high reliability and reasonably stable costs to the consumers. This article frames the history of the electric industry with the attempts to deregulate and provide for the benefits of a competitive environment into the 21st century.

INTRODUCTION

Accessible and reliable energy—in particular, electricity—is one of the driving forces of a successful economy. Since the early stages of the industrial revolution in the United States, abundant energy has been a cornerstone of domestic economic growth. Initially this growth was fueled by the creation of an inconsistent and unregulated utility system. The system energy supply business gave rise to the unregulated monopoly structures of the 1920s, which created ineffective and anticapitalistic business structures. These aggressive monopolies forced the regulated energy supply structure, encapsulated by the Public Utility Holding Act (PUCA) of 1935, and set in place the framework for today's regulated utility companies with the means of delivery and regulation of electric supply. For over 40 years, the resulting system helped build a foundation of manufacturing, service, and technology achievement. However, as with any system, changes are inevitable and the business of energy is no exception. Today, some customers vie for choice in electric suppliers and the potential for lower energy costs, but the structures of old do not allow for these opportunities. Starting with the Public Utility Regulatory Policy Act (PURPA) of 1978 and continuing with the Energy Policy Act of 1992, customer choice at the wholesale and the retail level were viable options but not without some major challenges. In the next few pages, the origins of competition in the electric industry to today will be traced and some of the results and current activities will be reviewed. Customer choice is a concept that strikes a chord at the heart of American business, and that chord is competition. The final tally has not been written because there are examples of success and failure in the U.S. market; however, activity continues in attempting to find the balance between reliability, quality, cost, and adequate supply that will drive the next century of economic growth.

Keywords: Electric competition; Customer choice; Open access; Reduced costs; Reliability; Wheeling; Wholesale energy trading.

THE BEGINNINGS

Electric supply systems were born in the city that never sleeps, New York, when Thomas Edison installed the first generation and distribution system to serve customers in the metropolitan area searching for a way to use a new idea called an electric light. From these early halcyon days, competition was fierce and fully unstructured as the race was started to see who could serve the needs of an interested and power hungry populace. This race gave way to multiple providers in close proximity and required oversight by government. As early as 1905, the city of Chicago had granted 21 municipal franchises to provide electricity including three franchises that covered the entire city.[1] However, this system of supply had too many choices with too few rules, which created the first natural monopolies based on economies of scale. The 1920s yielded the unregulated electric monopoly with stable costs but the need for growth beyond affluent customers, which were the primary customers of the time. In Pennsylvania, the state government proposed centralization of electric generation and a transmission system to connect existing customers and new customers and this was decried as un-American. This initial step was the first movement towards centralized generation with associated transmission system to deliver the fuel to power America's ascendancy to the top of the industrial peak. By 1932, nearly all investor owned utilities (IOU) moved towards the same centralized generation and transmission model proposed in Pennsylvania, but in an ominous turn, 50% of these IOUs were controlled by three companies. This ownership structure created a worry that unregulated monopolies would be able to control the electric system without any limitations. With unfettered control on the horizon, the federal government passed the Public Utilities Holding Act of 1935 (PUCHA) to limit electric monopolies and offer a framework for a structured generation and transmission system to supply electricity to a growing nation. This system and its evolution served well for nearly 40 years until the 1973 oil crisis shocked the nation into a new era of energy awareness and regulation.

GROWING PAINS

1973 was a year of long lines at gas stations and a collective shock to the economy based on the new limitations of energy. With this painful ordeal seemingly without solution, politicians and regulators alike searched for a way to help increase the opportunity for energy production from nontraditional central power stations. This "out of the box" thought gave birth to the Public Utilities Regulatory Policy Act of 1978 (PURPA), which had many components but for this discussion yielded the mainstreaming of cogeneration and of nonutility generation (power production) and direct interconnection if specific requirements were met by the energy producer. The two key components of PURPA in this regard were (1) that certain generators of electricity would be able to produce power and not fall under the utility regulation formats and (2) that there is a requirement of the local regulated utility to purchase excess generation from the nonutility generator at the utility's avoided cost.[2] This requirement opened the crack in the door for third-party generator interconnection to the electric grid and sale/purchase of this power. PURPAs intent was very clear, but the results were mixed. Because the state utility regulators oversaw the implementation of the program, calculations and processes surrounding avoided costs, interconnection standards, and local regulations created a patchwork of activity in the Unites States. California and New York were among the most ardent supporters of PURPA and the implementation of programs, but even today the rules in place for PURPA have directly impacted the efforts to restructure the electric markets.

In 1992, after many years of activity surrounding deregulation of the interstate natural gas supply industry, the Energy Policy Act of 1992 (EPAct '92) was passed to create the advent of electric market restructuring on a large scale. Based on the lesson learned from the natural gas market unbundling, the Federal Energy Regulatory Commission (FERC) provided the option of "wheeling" or open access across transmission systems based on the petition of buyers or generators of electricity. The effect of EPAct '92 was the initial opening of the wholesale transmission market, public information regarding transmission facilities, and expansion of the exemptions from utility regulation granted to certain generators started with PURPA. The main limitation of this law was that FERC was unable to extend open access to "retail" customers, but this right was reserved for the states to address on an individual basis.[3] These laws and events set into motion the start of the race for deregulation of the last true monopoly in American business.

EVENTS AFTER PASSING THE LAWS

After the passage of EPAct '92, the world of electricity was facing a new frontier. Wholesale energy trading was now an approved business and the race was on to see exactly what kind of business could be developed. As with any new business, the rules did not cover every possible permutation or scenario and so some of the practices had to be determined through experience. FERC used Order 888 to create the framework for movement of energy over the national electric grid in a way that promoted competition between generators to supply wholesale electric customers (for this discussion, wholesale energy customer includes regulated utilities, municipal utilities, rural electric cooperatives, and the like) based on more capitalistic principles. Order 888 (1996) required that all transmission owners and operators file tariffs stating rates, terms, and conditions for moving power over their transmission lines. Additionally, the order required that the transmission owners and operators charge comparative rates to what they would charge their utility or affiliates, thus offering a level playing field for moving power from one side of the transmission system to the other at standard prices.

Order 888 also set the foundation for the development of regional transmission operators (RTO) to control flow and access to the system. Prior to Order 888, the United States electric system was arranged into three major interconnections (East, West, and the majority of Texas, known as ERCOT). After the blackouts of 1965, the North American Electric Reliability Council (NERC) was created with ten regions dividing the country to address the causes of the blackout and to help prevent future occurrences. The regions developed operating, design, and communication procedures between member utilities and fostered cooperation between utilities to maintain electric grid reliability. These regions were created for traditional utilities and their operation and helped to provide processes and standards of support among utilities using the transmission grid. Under these rules and procedures, each utility owned, maintained, and controlled transmission systems in their franchise area or from their generation source to the load. However, the arrangement did not adequately address nonutility generators, which gave rise to the need for RTOs to help maintain reliability working with the existing utilities and enabling nonutility generators to have access to provide their product to the grid. RTOs would now be the hub for access and delivery of electricity to wholesale customers. The theory of the RTO was that pricing, maintenance, scheduling, development, expansion, and access would be most effectively controlled by one entity rather than allow for the multiple structures of franchise utility controls, as historically was in place. There was also a desire to prevent discriminatory practices to prevent nonutility generators from using the grid to supply alternatives to wholesale customers.

Is the RTO concept successful? The final tally has yet to be determined simply based on the fluid nature of the system. What was developed in the late 1990s is not the same system seen today based on the changes in regulations to adapt to business practices and precedents developed through regulator rulings on disputes and contentions. What is available for review is a comparison to the original goals of EPAct and Order 888, which included increased access to transmission systems by nontraditional generators, transparency in costs, nonpreferential treatment in transmission system use, and maintenance of existing reliability of service. With the exception of price spikes in the Northeast and the Midwest based on market forces and one reliability failure in 2003 for the Northeast, it appears that the wholesale side of electric deregulation has been mostly successful to this point based on these limited benchmarks. Is it a perfect system for "open" access and "full competition"? Not yet. There is much work remaining to make the process more effective and provide increased access, reliability, and control without dramatically increasing costs for wholesale and, in turn, retail consumers. Some of the major issues remaining include interconnection standards and processes, transmission planning, and capacity limitations.

WHOLESALE REASONING

Why was there a push to open the wholesale markets? Was the existing system broken? Were customers clamoring for full choice? These seem like simple questions, but in reality, the answers are very complex. The existing system was not broken. Utilities provided reliable and consistent energy supplies to wholesale and retail customers. However, the deregulation of other industries including natural gas, trucking, airlines, and telecommunications fully placed the spotlight on the one major remaining regulated industry. Additionally, international electric industry restructuring, especially in the United Kingdom, offered apparent support for the economic benefits of open markets. There were significant pricing disparities between the coastal regions and the central United States, and conventional wisdom held that a competitive market was, in the long run, the most efficient model for the consumer. These thoughts and others pushed regulators, lawmakers, and customer towards the opening of the markets.

At the time of PURPA in 1978, the country's leaders were searching for ways to reduce the cost of living and operating in the United States. Faced with increasing interest rates, a deepening recession, and a general increase in the cost of goods or services, politicians, regulators, and citizens were looking for ways to reduce costs (and lessen dependence on foreign energy sources—oil for starters). Increasing energy production from sources using combined heat and power looked like an ideal solution to address energy costs while allowing access to the electric grid and not increasing utility regulation or generating companies. It certainly seemed like a perfect first start. However, as with any first concept, there were many hurdles. What was initiated in 1978 needed many years and many lawsuits to find a balance when national events once again pushed legislators to action. This time, the events surrounded the turmoil in the Middle East via armed conflict. Once again, the focus was on the United States' apparent economic exposure to imported crude oil. This coupled with a heightened urgency to address environmental concerns led to the passage of the Clean Air Act Amendments of 1990 and the Energy Policy Act of 1992. Both of these provided an impact to the energy markets: the Clean Air Act started to address emission issues while the Energy Policy Act attempted to build on PURPA and natural gas deregulation.

However, what the regulation or deregulation of the wholesale power market did not address was the creation of a liquid market structure. It was anticipated that the "market" would create a clearinghouse and develop liquidity of price, much like commodity markets for natural gas. The acceptance of electricity as a trade-worthy commodity like oil, wheat, cotton, and natural gas by the financial markets was and is a cornerstone of a deregulated marketplace. In effect, if standard financial tools and techniques are available for large-scale energy trading, the market of supply and demand would set the price of the commodity based on pure economic principles. As with any new commodity market, it takes some time for experience and the application of hedging, derivatives, and other generally accepted trading tools. These tools were needed to help contract negotiations between suppliers and consumers to help provide some risk mitigation. As much as farmers who grow corn can use the financial markets to bracket their financial risk and improve their returns, so too would a municipality or large energy purchaser be able to do the same when buying electricity from suppliers in adjoining regions. The only problem was that electricity did not respond like other established trading commodities. With the inability to store electricity and inconsistent rules across varying transmission systems, there was great potential for supply and demand imbalances. The difficulties with trading electricity became apparent in the summer of 1998 when prices in the Midwest skyrocketed to more than $10,000 per megawatt. As the prices moved higher, contractual defaults started to increase. Suppliers and users were not ready for this type of fast price movement after the years of steady wholesale supply and contractual prices. Other regions were also affected with similar supply and demand imbalance events, and some of these, such as the activities in California, have created court actions to determine if markets were artificially manipulated.

Clearly the wholesale market for electricity moved from infancy to adolescence in a big hurry with all of the usual "heartbreak" one would associate with this growth. On the wholesale side of deregulation, the march continues forward in strengthening RTOs and developing more liquidity in the financial markets for electricity. Owners and operators of transmission systems and system traffic controllers are adapting to the speed and technical concerns to balance supply and demand in an open market. The complexities of a deregulated wholesale electric market have raised the bar on the need for accurate information flow, speed of response, flexibility of resources, and a host of other concerns. In general, the opening of markets has seen an increase in participation via nonregulated generators. The largest question still remains—is the deregulated marketplace better than the regulated version that was replaced? At this time, the opinions on this question remain split and are fully dependent upon issues of personal concern including reliability, delivered cost, access, environmental impacts, economic impacts, security, and many more. Not since telephone deregulation have more people in all stages of the value chain felt an impact of the change of an industry's structure, and the changes will continue as the experience grows. Only time will help to determine the success or failure of wholesale electric deregulation. A return to the previous system seems very unlikely and nearly a physical improbability.

RETAIL CONSTRUCTION

With a wholesale deregulated market driven by federal regulations and laws, the call for retail deregulation of electricity was not far behind (some might even say that the two calls were nearly simultaneous). The initial structure of PURPA and EPAct '92 addressed the concerns of large generators and cogeneration facilities. These large entities had and now have a greater ability of electric self-determination than a small business owner or a residential consumer. This is where the theory of retail deregulation tries to match consumer choice with the end consumer. There are many arguments advanced for deregulation of electricity to the home meter, including lower delivered costs, enhanced customer service, increased quality, faster development of new associated products and services, the reduction in power of the incumbent monopoly, and much more. The reality of course is much different than the theory.

The opening of wholesale markets was dictated by FERC and federal regulations, however, these same regulators and legislators decided not to address the retail customer due to large disparities in geography, existing regulation infrastructure, complexity of rule-making, and the rights of states to make and enforce rules pertaining to their citizens. With these thoughts and the need to focus on wholesale market structure, the federal regulators provided each state the flexibility to address retail service as it deemed. Reasoning that the existing regulatory structure would best know how and when to provide retail electric choice, federal regulators opted to concentrate on transmission and generation rules.[4] What this created was the potential for 50 different scenarios of retail electric competition with varied rules for access, service, reliability, and business practices.

In reality, the resulting "void" of federal mandates on retail competition provided for local choice in participating in the deregulation of electric suppliers. Debates of the issue were very strong in most all state regulatory and legislative arenas, with leadership in opening retail markets to electric competition developing in the coastal states. California, Rhode Island, Massachusetts, and New York were among the first states to open the retail market for electric competition in 1998. Citing high retail energy costs and the potential for dramatic reductions in electricity costs for consumers, markets were opened to energy suppliers. In the ensuing several years, other states followed with regulatory, legislative, and combination open-access rules for the retail electric markets. All total, 18 states have adopted retail electricity restructuring initiatives (California suspended their activities due to major issues in 2003) and another two states have large customer open access.[5] However, more than half of the states have opted to delay or not take any major action in the retail deregulation of electricity. If the opportunities are so good, why did not all states open their electric markets?

This is a large and complex question, but looking at some of the major concerns helps to shed light on decisions that dramatically affect the daily lives of individual and business consumers. Begin by seeking to find the driving factors of electric deregulation in the areas where these actions have taken place. Initial analysis provided an indication that customers could save between 5 and up to 40% in electricity costs if retail open access was available (these price reductions did not account for stranded asset cost recovery). Anecdotal discussions regarding the increased advance of new generation technology, the decreased environmental impact, and the improved customer service were also noted as benefits of customer choice in electricity. However, these points lacked validation from existing electric restructuring

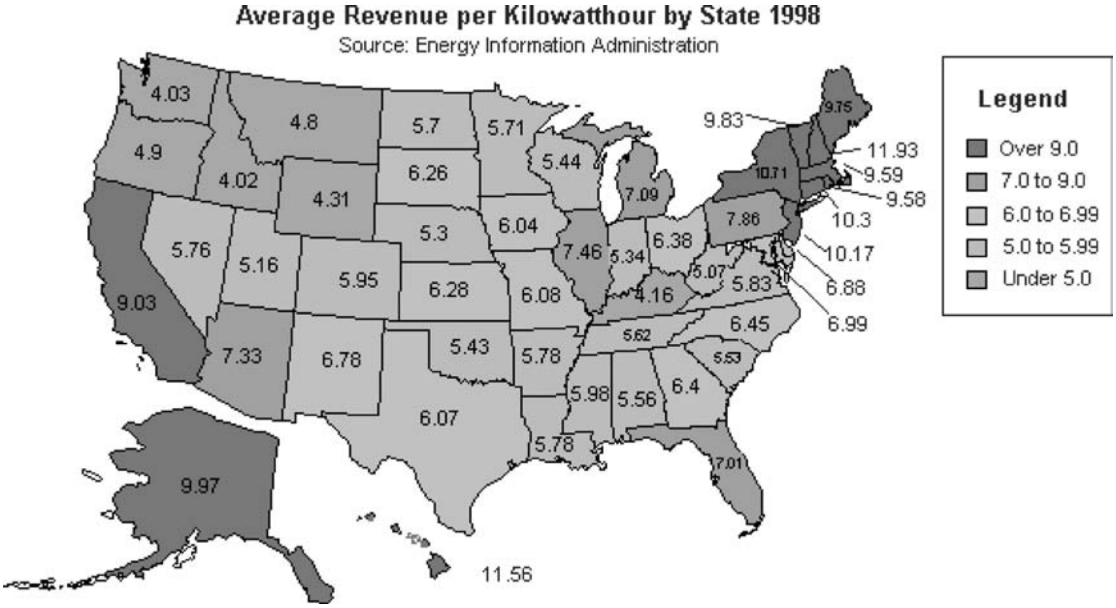


Fig. 1 Average revenue per kilowatthour by state 1998.
Source: From Energy Information Administration, 1998.

activities and appeared to have been interpolated or extracted from other industry's experiences with increased competition. So the remaining (and most prominent) driver of issues in United States business is economic improvement or lower consumer costs. Because this text is limited on space, there will only be a highlighting of each group's concerns, but this should help to define the overall image.

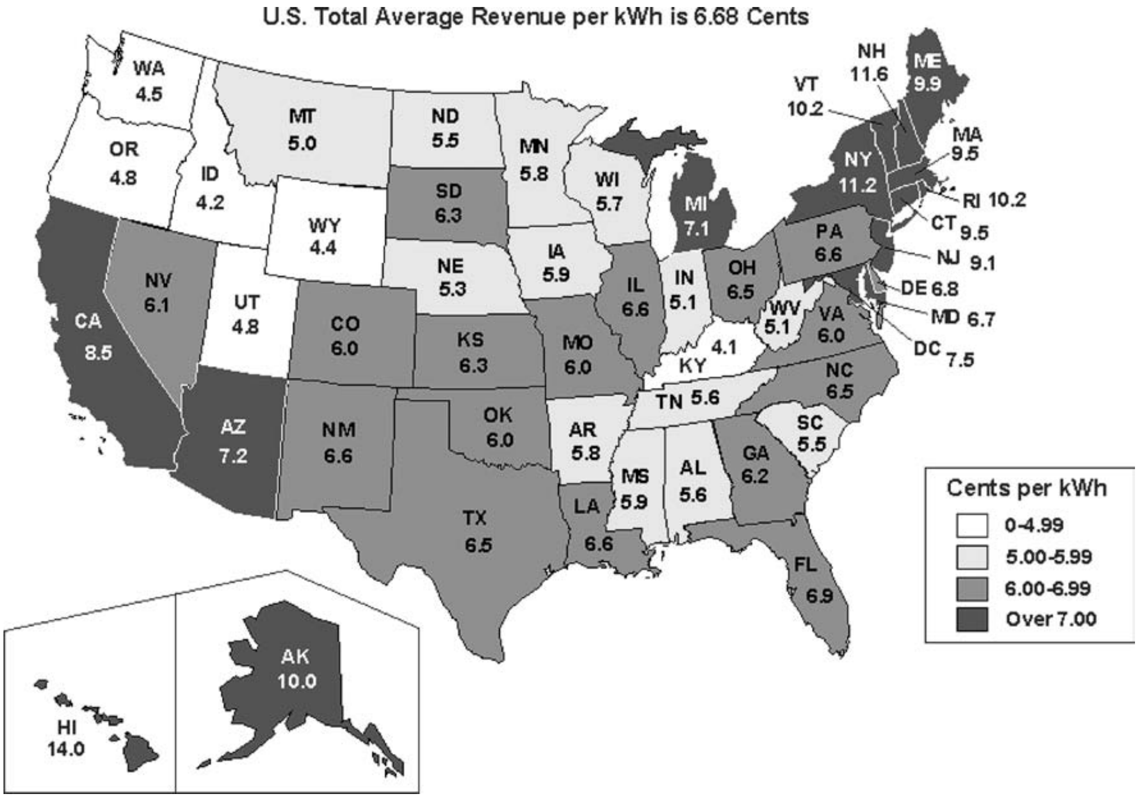


Fig. 2 U.S. total average revenue per kWh is 6.68 Cents.
Source: From Courtesy of Energy Information Administration, 2000.

Elec–Elec

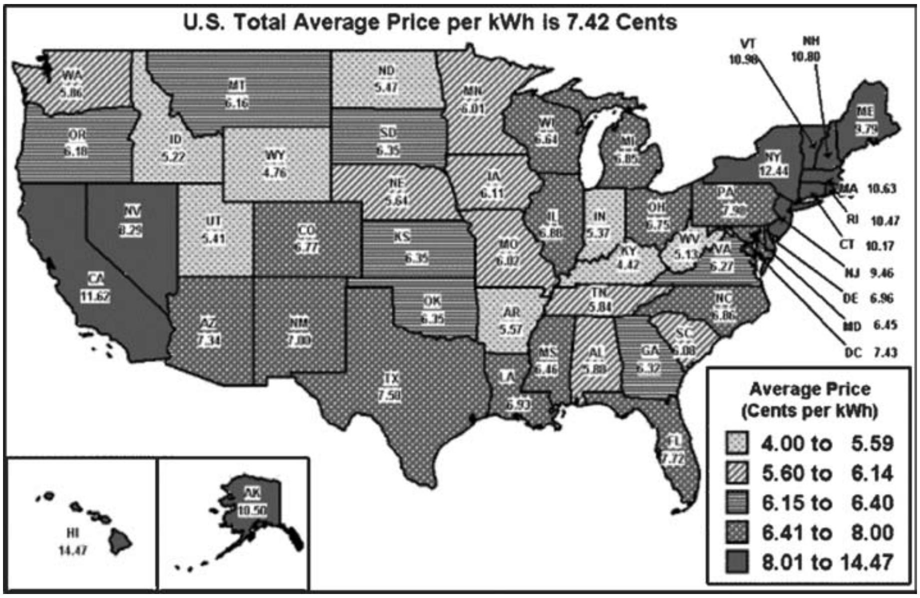


Fig. 3 U.S. total average price per kWh is 7.42 Cents.
Source: From Courtesy of Energy Information Administration, 2003.

THE CONSUMER'S POINT OF VIEW

The consumers of electricity are usually classified into residential, commercial, and industrial users. Of these, typically the industrial users are considered the largest consumers of electricity and thus they are the most prominent players in the economic situation. However, because electricity is also a regulatory/legislative issue, the residential and commercial interests are important in economic and political terms. With this in mind, one can generalize that industrial, commercial, and business clients are more apt to pursue cost reductions in energy to strengthen their bottom lines—they would view energy as any other commodity in their supply chain and use political pressure to advocate their business interests. From a residential economic perspective, lower costs of electricity would create availability of funds to meet basic needs or for discretionary spending and this group would use the court of public opinion to help open access. The majority of the costs savings certainly favor the larger users based on bulk purchasing capabilities; however, small consumers could have the chance to bundle purchasing power in some manner (the Sam's Club effect), but their impacts would be less than that of a single or multiple large business user for economic purposes if simply viewed in the increased complexity of supplying such an aggregate grouping. Additional consumer arguments of increased reliability and improved service based on ability to change suppliers (open market theory) are also very prevalent but not necessarily applicable based on the experience of open markets.

THE REGULATOR/POLITICAL POINT OF VIEW

Tip O'Neil once stated that "All Politics is local" and this is very evident in the actions regarding deregulation of electricity. Federal rules and regulations focused on the large picture of regional transmission, wholesale transactions, large-scale electric generation, and utility interaction. With a commitment in Order 888 that retail or customer choice was the domain of the states, FERC provided a Pandora's box and a golden fleece to each state utility regulator and lawmaker at the same time. The regulators had (and still have) complete control to develop programs that are beneficial to the local population. However, this also left the option of having 50 different programs and with many utilities covering multiple states, the potential for confusion on a grand scale. So what are some of the driving forces on the "local" level? In this case, the driving forces include (but are not limited to) responding to the call from customers to create a "free" market place, attempting to lower energy costs, increasing options in providers, increasing service, allowing once regulated companies to compete without going bankrupt, retaining property taxes from utilities, increasing the economic development capabilities for their areas, retaining the local energy marketplace, and improving the opportunities for the environment. This is a large list of opposing forces, and based on the states, the results and the main focus vary greatly. For example, in California, the main focus was on lowering consumer costs and potentially reducing environmental impacts. In Texas, the programs focus on customer choice and reliability, whereas in New York there is a focus to reduce customer costs (NY had the highest cost electricity provider in the continental United States) and offer options in the spirit of PURPA, of which they were a leading proponent. In other states such as Kentucky and Indiana, the focus is on maintaining the existing low electric rates and economic development opportunities for their communities.

So the regulators and legislators have multiple forces pulling them in many directions. Naturally, these opposing forces are not necessarily interested in the same outcome and the scales of balance on the issue move from one side to another, fully depending upon the individual regulators or political perspective. Looking at the various states and their positions regarding deregulation as well as the reviews of their implementations, one finds many interesting stories. For instance, California—which led the fray into deregulation under the concepts of lower electricity costs, increased open market activity, and reduced long-term environmental impacts—suspended their retail program in 2002 based on the results of the program, which included near bankrupt utilities, massive rolling blackouts, increased wholesale costs, and the like. Not all of these results stemmed from the regulators and political actions to create a deregulated marketplace, but the combination of the type of deregulation, market forces, market manipulation, and increased energy consumption helped to develop a very unstable system which tipped towards collapse. California is attempting to learn from their experience and at present is aggressively pursuing both supply and demand side activities to create balance, and perhaps at some point they might be able to advance their efforts to create an open market for retail customers as well as what exists for wholesale companies. In other states such as Pennsylvania, electric costs before deregulation (1996) and after deregulation have remained stable with reduction in cost for industrial customers of slightly more than 5% (based on EIA data comparing 1990–1995 vs 1996–2003 FSP costs). For Pennsylvania consumers, the regulators and legislators developed a program that appears to work and as the market matures, more benefits may be seen (Figs. 1–3). The following charts illustrate average costs for electricity in the United States. However, changes in costs cannot be solely attributed to deregulation due to the immense complexities of the mix of items that develop an average cost.

THE UTILITY POINT OF VIEW

Deregulation of retail electricity service was a shocking blow to the electric utility industry in that they had approached the business with a long-term (30-year view) based on the systems in place through PUCHA. PURPA created several new issues, such as qualified facility energy production, nonutility interconnects for wholesale power sales, and so on. The progression of FERC Order 888 and the opening of the wholesale market caused an increase in the angst at the utility company. In effect, FERC 888 created a new way to address wholesale concerns from a physical and financial approach for the utilities. Physically, the utilities were used to move power from point one to point two based on some economic activity but mostly based on reliability needs as developed by the NERC control program. Now, the same transmission system was to be used for economic dispatch. System models that were built for reliability now needed to act as a highway. To further the complications, new nonutility generators (NUG) were looking to establish facilities to sell power to areas in which they could potentially create a profit for their shareholders.

The physical challenges were intertwined with financial challenges and the existing structure. Utilities were accustomed to planning on long-term cycles and power plants, transmission systems, distribution systems, staffing, and so forth were determined through a rate-based process using integrated system planning. The infrastructure needs developed from this type of analysis and the allowable rates of returns guided investment decision for regulated utilities. NUGs did not develop or perform business in this way, which set the stage for competition

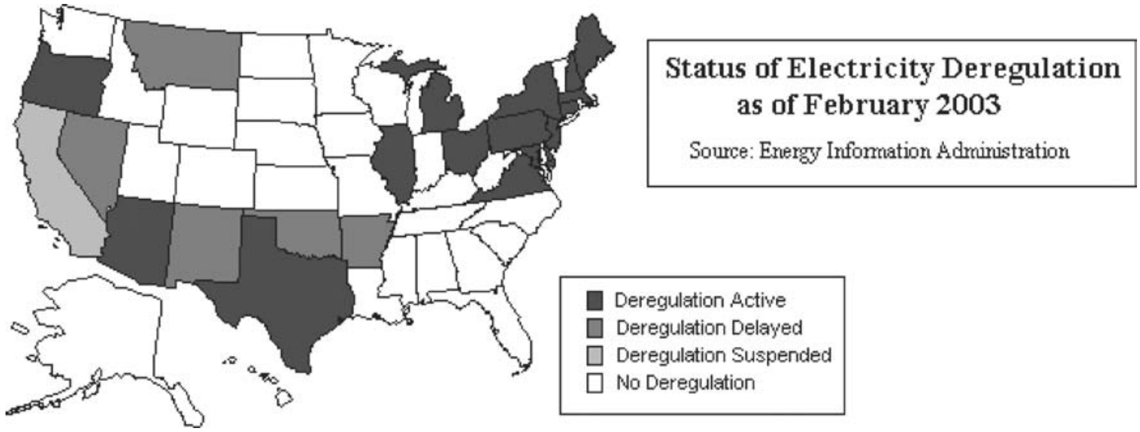


Fig. 4 Status of electricity deregulation as of February 2003. Source: From Energy Information Administration.

among various business models with varying rules. These rule variations lead to the discussion of stranded assets (those assets built by regulated utilities in past years with the understanding the assets would be recovered or paid by the customers over time). When deregulation became a distinct possibility, utilities naturally were worried that investments in infrastructure prudently made under the "old" rules needed to be addressed if customers began to select new suppliers. Some of these concerns were addressed on the wholesale side through FERC but because the majority of issues arose on the state level, each state had to determine the best way to balance the stranded asset issue without front loading prices in a deregulated market. This would in effect increase consumer prices in the short run if stranded assets would be recovered in a short period of time. Other issues concerned billing of customers as well as providing service to credit risk consumers. Electric utilities have a requirement to serve customers with highly regulated disconnect or refusal of service processes. These rules would not necessarily be applied to unregulated competitors and thus they would conceivably leave the highest risk customers with the incumbent utility.

These were just some of the concerns of each stakeholder in the deregulation story. Currently, there are 16 states with active programs and another five states that have delayed implementation but are ready to start once specific concerns are addressed. California has suspended their deregulation experiment based on the many issues and the existing public debt created by the original program. Their actions in attacking the demand side portion of the energy will solve one of the concerns to help balance the energy equation. The initial deregulation design did not offer any consumer triggers to reduce consumption based on price signals or supply/demand imbalances, which helped to drive up costs and, in part, increase the probability of demand out-stripping supply. In this area of demand control, California is showing national leadership and will increase the likelihood of future resumption of deregulation. The majority of states have decided to continue to study deregulation or take no action. Their existing price and supply structure shows little or no consumer benefits to pursue open markets. As energy prices increase and supply tightens, even these markets may open to a more competitive environment, but the shock of events surrounding other unsuccessful actions will increase the threshold for action.

CONCLUSION

Deregulation of electricity comes in wholesale and retail flavors with distinctly different goals and results. (Fig. 4) Wholesale open access started with PURPA in 1978 and marched forward with FERC 888 and EPAct 1992 and continues through state retail access programs. So is deregulation of electric markets beneficial to consumers? It depends on who is asked and their perspective of the situation. In most cases (based on information from EIA), the consumers have seen some benefits with increased complexity of the offering. This appears equivalent to the scenario of long distance telephone competition in which consumers are inundated with varying offers. Through scrutiny by state regulators, these electric offerings are more controlled and thus not as likely for confusion. The story of electric deregulation is far from complete and many chapters are yet to be written. Lessons learned from unsuccessful as well as successful programs will help guide regulators and legislators to find balance between consumers, utilities, and electric providers. Based on other previously competitively regulated industries, competition has proven some benefits. The stakes with electricity are very high because this country is so dependent on electricity and this major fact warrants extreme caution in program design and implementation. So far, caution has been applied and the future for competitive electricity is waiting for a technology break-through to help ease the transition. Only time will tell.

REFERENCES

1. Christensen, P.C. *Retail Wheeling—A Guide for End-Users*, 3rd Ed.; Pennwell Publishing: Tulsa, OK, 1998.
2. P.L. 95–617, Public Utilities Regulatory Policy Act of 1978, www.thomas.loc.gov.
3. P.L. 102–486, Energy Policy Act of 1992, www.thomas.loc.gov.
4. *Commission Orders Sweeping Changes for Electric Utility Industry, Requires Wholesale Market to Open to Competition*, Convergence Research, 1996, www.converger.com/FERC-NOPR/888_889.htm.
5. Status of Electricity Competition in the States, Edison Electric Institute, 2003, www.eei.org/industry_issues/electricity_policy/state_and_local_policies/state_restructuring_and_regulatory_policy/Competition_states_map.pdf.

BIBLIOGRAPHY

1. *Electricity Prices in a Competitive Environment: Marginal Cost Pricing of Generation Services and Financial Status of Electric Utilities. A Preliminary Analysis Through 2015*, Energy Information Administration, 1997, DOE/EIA—0614.

Elec–Elec

Electricity Enterprise: U.S., Past and Present

Kurt E. Yeager
Galvin Electricity Initiative, Palo Alto, California, U.S.A.

Abstract
This article provides a brief historical synopsis of the development of the electricity enterprise in the United States. The end of this article includes a reference list of more detailed accounts of the electricity enterprise, on which this summary is based.

INTRODUCTION

The organization of this synopsis corresponds with the stages in the Electricity Sector Life-Cycle shown in Fig. 1. The first three stages of the life-cycle are discussed in this article. A second article discusses the future of the U.S. Electricity Enterprise.

Electricity now powers American life to an unprecedented degree. In 1900, electricity's share of the nation's energy use was negligible. That share has now risen to nearly 40%. Electricity is produced from fuels through a costly conversion process so that its price per thermal unit has always been higher than that of the fuels themselves. So, something other than cost must account for the sustained growth in electricity's market share. Simply put, electricity can be used in ways that no other energy form can. Technological progress over the past century has led to radically improved ways of organizing productive activities as well as new products and new techniques of production, all of which have been heavily dependent on electricity. As a result, electricity has become the lifeblood of the nation's prosperity and quality of life. In fact, the U.S. National Academy of Engineers declared that "the vast networks of electrification are the greatest engineering achievement of the 20th Century."

Fig. 2 summarizes the historical trends in the energy sources for U.S. electricity generation. Coal is notable in its persistence of as the dominant fuel.

Electricity, despite its mystery and complexity, is simply the movement of electrons. Each of these tiny sub-atomic particles travels only a short distance as it displaces another electron around a circuit, but this transfer occurs at the speed of light (186,000 mi per second). This invisible wonder occurs virtually everywhere in nature. For example, it transmits signals from our brains to contract our muscles. What's relatively new is our ability to put electricity to work lighting and powering our world.

For example, steady advances during the course of the 20th century improved electric lighting efficiencies a great deal. Looking forward, full-spectrum light-emitting diodes (LEDs) may increase the efficiency of U.S. lighting by 50% within 50 years. At the same time, electric motors have revolutionized manufacturing through unprecedented gains in the reliability and localized control of power vis-à-vis steam engines. By 1932, electric motors provided over 70% of all installed mechanical power in U.S. industries. The proliferation of household appliances has also been primarily due to the use of small electric motors. Today, the ubiquitous electric motor in all its forms and sizes consumes two-thirds of all U.S. electricity production.

Electricity is indeed a superior energy form; however it is not a tangible substance, but rather a physical effect occurring throughout the wires that conduct it. Electricity must be produced and consumed in absolutely instantaneous balance and it can't be easily stored. Its delivery, therefore, today requires the ultimate just-in-time enterprise that balances supply and demand at literally the speed of light. Yet the status quo suffers numerous shortcomings. Efficiency, for instance, has not increased since the late 1950s, and U.S. generators throw away more energy than Japan consumes. Unreliable power—the result of blackouts or even just momentary surges and sags—costs America more than $100 billion annually. This is equivalent to about a 50¢ surcharge on every dollar of electricity purchased by consumers.

Moreover, the U.S. bulk electricity infrastructure is aging and becoming obsolescent. The average generating plant was built in 1964 using 1950s technology, whereas factories that construct computers have been replaced and updated five times over the same period. Today's high-voltage transmission lines were designed before planners ever imagined that enormous quantities of electricity would be sold across state lines in competitive transactions. Consequently, the wires are often overloaded and subject to blackouts. Yet demand is increasing at twice the rate of capacity expansion. Finally, the local distribution

Keywords: Electric enterprise; Restructuring; Electric system infrastructure; Electric system planning; Electric system history; Electrification; Electric transmission system planning; Grid system; Electric system transformation.

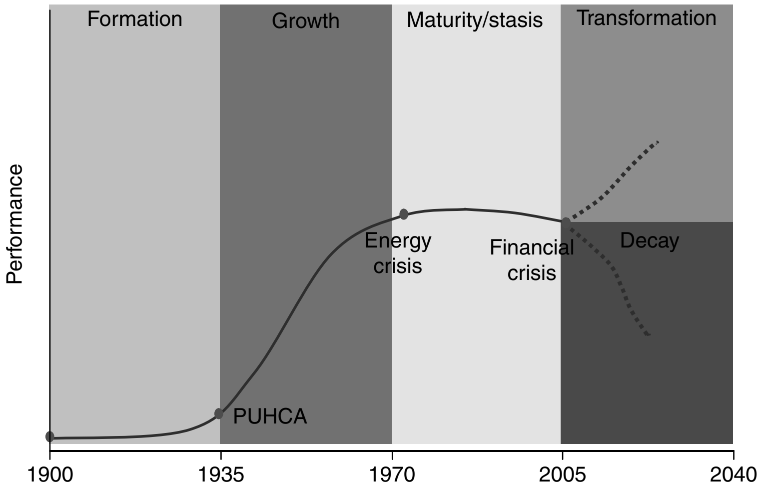


Fig. 1 Electricity sector life-cycle—a fork in the road.

systems that connect the power supply to each consumer are effectively a last bastion of analog, electromechanically controlled industry. This is a particularly notable paradox given the fact that the nation's electricity supply system powers the digital revolution on which much of the current and future value depends. Keeping the lights on 99.97% of the time is simply not good enough. That still means the average consumer doesn't have power for 2.5 h a year. In today's impatient, increasingly computerized world that is more than just a nuisance.

In spite of these deficiencies, the traditional producers and deliverers of electricity—the nation's electric utilities—hold assets exceeding $600 billion; with 70% invested in power plants, 20% in distribution facilities, and 10% in transmission. They form one of the largest industries in the United States—roughly twice the size of telecommunications and nearly 30% larger than the U.S. automobile industry in terms of annual sales revenues. The Achilles' heel here is the fact that supplying electricity is also extremely capital intensive, requiring far more investment per unit of revenue than the average manufacturing industry. This investment challenge is further intensified by the fact that the U.S. electricity enterprise is made up of over 5000 commercial entities, both public and private. The largest individual corporate market cap in the enterprise today is that of Exelon at $32 billion. This compares with Exxon-Mobil, for example, with a market cap of $365 billion. In fact, only about 17 electric utilities have market equity value greater than $10 billion. As a result, the decision to invest the billion or more dollars needed to construct a major new power plant or power delivery (T&D) line is effectively an uncertainty-laden, long-term, "all or nothing" decision that is constantly avoided by most electric enterprise corporations today.

Meanwhile, the market for portable electric devices continues to grow dramatically from the traditional flashlights and auto ignitions to a diverse array of

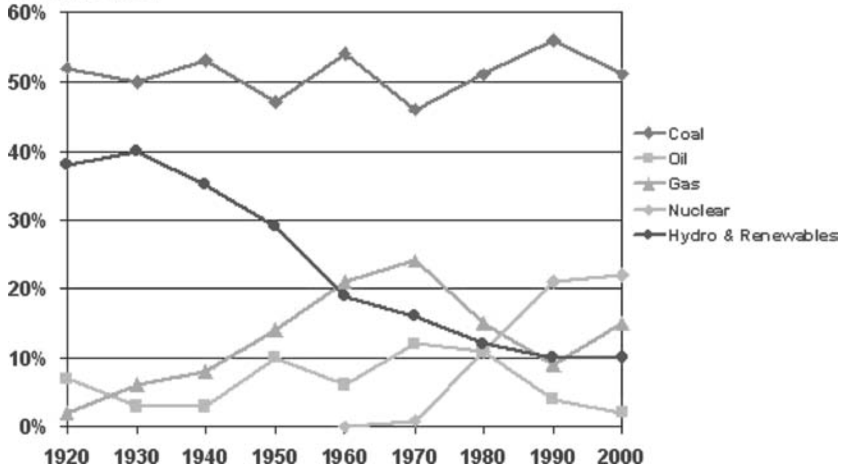


Fig. 2 Sources of energy for U.S. electricity generation.

computers, communications and entertainment products, cordless tools, medical devices, military products, etc. This innovative diversity has been accomplished by exploiting the synergy between the products themselves, the electricity storage devices they employ—including batteries, ultracapacitors, and fuel cells—and the power-management systems that charge these storage devices. Today, the global portable electricity storage market is about $50 billion per year, of which, $5 billion is allocated to rechargeable (secondary) batteries. Demand growth in this market is estimated at 6% annually, compared to grid-supplied electricity demand growth of about 1.5% per year.

A new generation of energy-hungry devices, such as digital cameras, camera phones, and high-performance portable computing devices, is expected to continue to drive this rapid growth in portable electrification. Notably, the kWh-equivalent price of portable electricity can be as much as $100, compared to about 10 cents for grid-supplied power. This is one indication of the potential for highly flexible, individualized electricity services to enhance electricity's value beyond the commoditized energy value proposition and business model of the traditional electricity enterprise.

The history of this magical energy form, electricity, provides keen perspective on the real-world interplay among technical progress, business struggles, and political debates. Electricity's ongoing evolution also suggests how the potential for renewed electricity-based innovation could curtail pollution and spur a wide array of electro-technical advances, while continuing to improve the U.S. quality of life, and that of billions more worldwide.

> Electrification is not an implacable force moving through history, but a social process that varies from one time period to another and from one culture to another. In the United States electrification was not a thing that came from outside society and had an impact; rather, it was an internal development shaped by its social context.

Table 1A and B summarize some recent basic statistics for the U.S. electricity enterprise.

FORMATION

Neither electricity nor electric lighting began with Edison. In 1808, Sir Humphrey Davey sent a battery-powered electric current between two carbon rods to produce an arc of light. In 1831, Michael Faraday invented the dynamo, which, when turned by a steam engine, supplied a cheap electric current by means of electromagnetic induction. Davey's arc lamp was used in a production of the Paris Opera in 1844 and was part of the Philadelphia Exposition in 1876. The development of the high vacuum Crookes tube served to rekindle interest in incandescent electric lamps that had begun in the 1820s.

Edison, with a characteristic vision that distinguished him from his competitors, worked not only on the incandescent lamp but on the entire system that powers the lamp. "The same wire that brings the light will also bring power and heat—with the power you can run an elevator, a sewing machine, or any mechanical contrivance, and by means of the heat you may cook your food." Scientists and rival investors predicted failure. In 1879, Edison had a working incandescent lamp and, within months, patented a direct current (DC) electric distribution system.

The next question was: who would buy the lights and equipment? Edison conceived the central power station that would distribute electricity through lines to the customers. On September 4, 1882, Edison's Pearl Street Station in New York City went into business to serve 85 customers with 400 lamps. This marked the beginning of the electric utility industry. Unfortunately, with Edison's low-voltage DC system, it was too expensive to distribute electricity more than a mile from the power plant. Transformers that could raise and lower voltage did not work with direct current.

In 1888, Nicola Tesla, who had been previously employed by Edison, announced his polyphase alternating current (AC) power system. That same year George Westinghouse bought the rights to Tesla's system. Westinghouse saw the potential for locating a central station at the source of water power or coal, shipping the power for great distances at high voltages, and then stepping down the power for distribution. But inventing a practical AC system created new problems. AC and DC systems could not be linked until Charles Bradley, another former Edison worker, invented the rotary converter in 1888 which converted DC to AC. Westinghouse also bought Bradley's idea.

The Westinghouse engineers developed a universal system in which the polyphase AC generator at the central station produced electricity that went to a local substation where it was transformed to the voltage required by the user. This system had many advantages. First was a realization of economies of scale in power generation. The second was the need for only one wiring grid. The third was that the generating stations could serve a wider area. The fourth was that the new system's productivity could benefit from load diversity; e.g., lighting in the evening, streetcars during rush hours, and factory motors during the periods in between. Interestingly, the introduction of practical electric streetcars in the late 1880s provided the major concentrated electricity demand that dramatically pushed the enterprise toward more powerful equipment and larger service areas favoring AC.

In little more than a decade, Edison had put more than a half century of research into practical applications, conceived and invented an entire industry, and then

Table 1A U.S. electricity industry statistics (2002)

Generating capacity and generators	Capacity (1000 MW)	Net generation (billion kWh)
Investor-owned utility	398	1740
Government and cooperatives	204	807
Non-utility (unregulated producers)	380	1294
Total	982	3841
Total net generation by fuel	**(Billion kWh)**	**Percentage**
Coal	1926	50
Nuclear	780	20
Natural gas	695	18
Hydroelectric and pumped storage	253	7
Fuel oil	92	2
Biomass	72	2
Geothermal	13	1
Wind	9	
Photovoltaic	1	
Total	3841	100
Electricity sales	**Customers (million)**	**Energy (billion kWh)**
Residential	115	1267
Commercial	15	1122
Industrial	0.6	973
Other	1	109
Total	132	3471

Revenues	Sales ($billion)	Percent total electricity energy (%)	Average/kWh (cents)
Residential	107	37	8.4
Commercial	89	32	7.9
Industrial	47	28	4.8
Other	7	3	6.6
Total	250		Average price 7.2

Financial	($Billion)
Total assets	598
Total operating revenues	250
Operating expenses	220
Operating income	30
Construction	25

Source: From Edison Electric Institute (see Ref. 1).

became a reactionary who threatened to stagnate the industry at in its primitive stage of development. In 1892, Edison's financier, J.P. Morgan, stepped in and forced a merger with Thompson-Houston and put their management in charge of Edison's General Electric Co. In 1896, General Electric and Westinghouse exchanged patents, a typical move in the age of trusts, so even General Electric used Westinghouse concepts. The age of Edison had ended.

By 1892, Samuel Insull of Chicago Edison, another former Edison associate, had formulated an understanding of the economics of the electric utility business that was

Table 1B U.S. consumption of electricity (2001)

A. Residential	Billion kWh	Percentage
Air conditioning	183	16
Refrigerators	156	14
Space heating	116	10
Water heating	104	9
Lighting	101	9
Ranges and ovens	80	7
Laundry	76	7
Color TV, VCR/DVD, Stereos	55	5
Freezers	39	3
Furnace fans	38	3
Dishwashers	29	2
Personal computers and communication	28	2
Pools and hot tubs	17	2
Other[a]	118	10
Total	1140	100
B. Commercial	**Billion kWh**	**Percentage**
Space cooling	288	26
Lighting	255	23
Office equipment/computing	200	18
Refrigeration	100	9
Ventilation	78	7
Space heating	56	5
Cooking	22	2
Water heating	11	1
Other	100	9
Total	1110	100
C. Manufacturing	**Billion kWh**	**Percentage**
Machine drive	512	53
Process heating	104	11
Electro-chemical	86	9
HVAC	82	9
Process cooling and refrigeration	63	7
Lighting	62	6
Other	61	6
Total	970	100

[a] Composed of about 15 additional consumption categories, each representing less than 1% of residential electricity consumption.
Source: From U.S. Energy Information Agency (EIA), Energy Consumption Survey.

sustained through most of the 20th century. When Insull took over the Chicago Edison Co. in that year, it was just one of 20 electric companies in the city. Although Chicago had a population of more than one million, only 5000 had electric lights. He vowed to serve the entire population. Insull and other leaders of the Association of Edison Illuminating Companies (AEIC) realized that the industry had high fixed costs because of the infrastructure investment needed. At the same time, the cost of operating the plants was fairly low. The question was how to translate that into profits, especially in an industry that had concluded it was selling a luxury item.

Insull began a sales campaign, cut prices as necessary to get customers, and wrote long-term contracts for large customers. He utilized a demand meter (invented in England) and set the price of electricity to cover both fixed and operating costs. Insull also concluded that profits were maximized by keeping the power plant running as much as possible to exploit the diversity of load. As a result, the U.S. led the world in the rates of electrification. Insull in Chicago sold more electricity per capita, ran larger power stations, kept the plants running longer each day, and charged customers less. Insull also discovered that there were clear advantages to tying together urban and rural loads. For example, Chicago had a winter peak and the farm towns a summer peak.

By 1911, thanks to the development of the ductile metal filament lamp, electric lighting ceased to be a luxury, manufacturers developed new uses (e.g., refrigerators and sewing machines), and the demand for electricity skyrocketed. Also, Charles Parsons had recognized the limits of the reciprocating steam engine in 1884, and developed the steam turbine that produced rotary motion directly as high-pressure steam pushed against blades attached to a shaft. This elegantly simple machine occupied one-tenth the space, and cost one-third as much as the reciprocating engine of equivalent capacity. By 1911, 12,000 kW turbine generators also became the norm. Thus, the keys to the success of the traditional, declining cost commodity, grow-and-build electric utility business model were established, i.e., economies of increasing scale, rapidly rising consumer demand, and consumer diversity for load stabilization and higher capacity factors.

However, during much of this era of rapid sales growth and technological progress, electric utilities managed to earn unspectacular profits. This was addressed by consolidating the over-fragmented industry into ever-larger holding companies. Centralized ownership served to facilitate raising money and engineering the best systems. Also non-utility, industrial, electricity generation declined from more than 50% of the U.S. total as late as 1914 to 20% by 1932. Although states regulated the operating subsidiaries that sold electricity, none regulated the holding companies. By 1932 the eight largest holding companies controlled 73% of the investor-owned electric businesses. The Insull empire alone operated in 32 states and controlled at least a half billion dollars in assets with an investment of only $27 million. As a result of these excesses committed, the electricity holding companies were condemned in the wake of the Depression, and controlling legislation was passed that created the present structure of the electric utility industry.

Under this legislation, interstate holding companies had to register with the SEC. This included any company that owned 10% or more of the voting securities of an electric or gas utility. The Act also broke up holding company systems that were not contiguous and eliminated intermediate holding companies from the financial structure.

Table 2 Electrification of the U.S. economy

	1902	1932
Percentage of population in electric-lighted dwellings	2	70
Percentage power in industry (horsepower equivalent)	5	73
Average power plant size (MW)	0.5	8.5
Electricity generation (10^9 kWh)	6	100
Residential service price (¢ per kWh—1983$)	~40	15

This Public Utility Holding Company Act of 1935 (PUHCA) also effectively marked the end of the formative period of the U.S. electricity enterprise. Table 2 summarizes the rapid progress of the electricity enterprise during this formation period.

GROWTH

In 1932, Franklin Roosevelt denounced the "Insull monstrosity" and proposed four Federal hydropower projects—The St. Lawrence, Muscle Shoals, Boulder Dam, and the Columbia. "Each of these in each of the four quarters of the United States will be forever a national yardstick to prevent extortion against the public and to encourage the wider use of that servant of the people—electric power." In the same general time frame, Lenin also underscored the universal impact of electricity by declaring that "Communism equals the Soviet power plus electrification."

Even during the Depression and through World War II, the U.S. electric utility industry continued to expand and to cut its costs. Government-supported entities, such as the Rural Electrification Administration, brought electricity to the farms. Although investor-owned utilities lost territory to governmentally owned (public power) utilities, the most significant change was the devolution of operating control from holding companies to locally operated utilities. These were incented to concentrate on customer service rather than on complex financial frameworks. Between 1935 and 1950, 759 companies were separated from holding company systems, and the number of registered holding companies declined from over 200 to 18.

During this period of industry consolidation and growth, utilities desired three features from new technology: reliability, greater power at lower costs, and higher thermal efficiency. As a result of this demanding market, manufacturers initially developed their new machines using a "design-by-experience" incremental technique. As with many other engineering endeavors then, the people who built these complex machines learned as they went along. This was reflected by steady increases

Elec–Elec

in steam pressure and temperature in boilers and generators, providing corresponding improvements in thermal efficiency. Steam temperature and pressure in 1903 were typically 530°F and 180 PSI respectively. By 1930, water-cooled furnace walls permitted the production of steam at 750°F and upto 1400 PSI. By 1960, these parameters had increased to 1000°F and 3000 PSI, turning water into dry, unsaturated, supercritical steam, effectively exploiting the full potential of the Rankine steam cycle.

Improvements in transmission systems also occurred incrementally during the power industry's first several decades. While comprising a relatively small portion of a power system's total capital cost, transmission systems nevertheless contributed significantly to providing lower costs and more reliable service. They did this by operating at ever-increasing voltages and by permitting interconnections among different power plants owned by contiguous power companies. Transmission voltage increased from 60,000 V in 1900 to 240,000 V in 1930 and upto 760,000 V by 1960. Increased voltage, like higher water pressure in a pipe, allows more electricity to pass through a transmission wire. Doubling the voltage, for example, increased a line's volt-ampere capacity by a factor of four. In short, the development of high-voltage transmission systems contributed as much to the steady increase in capacity of power production units as did advances in turbine speed or generator-cooling techniques.

U.S. energy consumption grew in lock-step with the economy after World War II, but electricity sales rose at double that rate until about 1970. The result, over the period of 1935–1970, was an 18-fold increase in electricity sales to end users with a corresponding 12-fold increase in electric utility revenues. This growth was stimulated by the dramatic and continuing drop in the real price of electricity, compared to other fuels. Much of the success in reducing costs was due to these continued improvements in the generating process and in higher voltages and longer distance transmissions, which together more than offset the impact of the break up of the holding companies. Over the post-war (1945–1965) period, the average size of a steam power plant rose fivefold, providing significant economy-of-scale advantages.

Efficiency improvements (heat-rate) did not, however, keep pace after the late 1950s, even with higher operating steam temperatures and pressures. The inherent limitations of the Rankine steam cycle coupled with metallurgical constraints caused this efficiency plateau. Electricity distribution system expense per customer year also increased over 100% during this period, from $8 to $17.

After World War II, the accelerated growth of the industry caused manufacturers to modify their incremental, design-by-experience approach to one of "design-by-extrapolation." This enabled manufacturers to produce larger technologies more rapidly. The push for larger unit sizes reflected the postwar U.S. economic prosperity and the introduction of major new electricity uses including air conditioning, electrical space-heating, and television. The all-electric home loomed large on the horizon. The biggest concerted promotional push began in 1956 with the "Live Better Electrically" campaign employing celebrities such as Ronald Reagan, on the heels of the very successful "Reddy Kilowatt" mascot for modern electric living.

While the best steam turbine generating units only improved in thermal efficiency from 32 to a 40% plateau in the postwar period, turbine unit sizes jumped from about 160 MW in 1945 to over 1000 MW in 1965 through the design-by-extrapolation approach. Correspondingly, new plant construction costs declined from $173/kW in 1950 to $101/kW in 1965. Between 1956 and 1970, utilities operated 58 fewer plants to produce 179% more electricity. As regulated monopolies, electric utilities could not compete with each other for market share, but competition existed during this period as engineer-managers strived for technical leadership among their peers. This type of competitive environment contributed to rapid technological advances and production efficiencies. Utility managers encouraged manufacturers to build more elegant technology so they could get credit for using the "best" machines. The risks to gain customized technological supremacy often meant an economic tradeoff, but this was a price readily paid in the 1950s and 1960s by utility managers who retained their engineering values and goals as they became leaders of large business enterprises.

During this period of rapid expansion and success, a third participant—in addition to electric utilities and manufacturers—played a largely invisible supporting role. This third party consisted of the state regulatory bodies, which performed two tasks relative to the electricity enterprise. First, they protected the public from abusive monopoly practices while assuring reasonably priced, reliable utility service. Second, they guaranteed the financial integrity of the utility companies. Conflicts rarely arose because utilities were steadily reducing their marginal costs of producing power, and they passed along some of these savings to consumers. Thus, few people complained about a service where declining costs countered the general trend toward cost-of-living increases. Regulatory actions also tended to reinforce the industry's grow-and-build strategy by permitting utilities to earn a return only on capital expenditures. This "social contract" served the industry and its stakeholders well for more than half a century providing a robust, state-of-the-art infrastructure. Although not articulated at the time, these stakeholders had forged an implicit consensus concerning the design, management, and regulation of a national technological system. As long as benefits continued to accrue to everyone, the consensus remained intact.

For electric utilities, this consolidation and growth period was, in summary, one of reorganization out of the holding companies, minimal need for rate relief, declining costs and prices, an average doubling in electricity demand

Table 3 Electricity enterprise growth

	1932	1950	1968
Ultimate customers (million)	24	43	70
Net generation (10^9 kWh)	100	389	1,436
Installed generating capacity (10^3 MW)	43	83	310
Average power plant size (MW)	8.5	18	85
Circuit miles of hi-voltage line[a] (10^3 mi)	NA	236	425
Residential service price (¢ per kWh—1983$)	15	10	7.1

[a] 22,000 V and above.

every decade, incentives to add to the rate base, satisfied customers and investors, and acceptable returns for owners. That environment of few operating problems and little need to question the prevailing regulatory structure left the electricity enterprise and its stakeholders unprepared to either anticipate, or respond quickly to, the challenges that rapidly followed.

Table 3 summarizes the progress of the electricity enterprise during this period of growth and consolidation.

MATURITY AND STASIS

By the mid 1960s, the electricity enterprise and its stakeholders were beginning to experience the first cracks in the traditional business model and its associated regulatory compact. The fundamental concepts of the enterprise began to be challenged and investment started to erode. The most notable initial event was the November 9, 1965 Northeast Blackout that spread over 80,000 mi^2 affecting 30 million people. This, and other outages that followed, forced utilities to redirect expenditures from building new facilities to improving the existing ones. Specifically, they had to upgrade the fragile transmission and distribution (T&D) system in order to handle larger power pools and more frequent sales among utilities. These new costs led to higher rates—for the first time, literally, in decades—and despite expensive public relations efforts, the public grew increasingly critical of utility monopolies.

1967 marked a second major turning point for the U.S. electricity enterprise—generation efficiency peaked. Rather than lower the average cost of electricity, a new steam power plant would henceforth increase it. Economies of scale ceased to apply (bigger was no longer necessarily better or cheaper) and continued expansion in the traditional manner no longer held the same consumer benefits. The grow-and-build strategy had seemingly reached the end of the line. A third turning point was Earth Day in 1970. This launched environmental activism and focused fresh attention on electric utilities, ultimately leading to further investment redirection for environmental control equipment, most notably for sulfur dioxide scrubbing on the industry's fleet of coal-fired power plants. The Clean Air Act of 1970 made environmental concerns an integral part of the utility planning process while planning for growth became more difficult.

The fourth major turning point event for the electricity enterprise was the Oil Embargo of 1973. OPEC's actions led to a rapid rise in the cost of all fuels, including coal. Accelerated inflation and interest rates resulting from the Vietnam War economy also led to higher borrowing rates for utilities. The sum of these turning point issues led to ever higher electricity prices, reducing the growth of U.S. electricity sales in 1974 for the first time since World War II. Consolidated Edison missed its dividend and utility stock prices fell by 36%, the greatest drop since the Depression.

In spite of these troubling events, the electricity enterprise's commitment to growth was slow to respond. In 1973 electric utilities issued $4.7 billion in new stock, almost seven times that sold by all U.S. manufacturing companies combined. Finally in 1975, capital expenditures declined for the first time since 1962. The traumatic decade of the 1970s concluded with perhaps the most strategically serious turning point issue of all—three mile Island. On March 28, 1979, a cooling system malfunction at the 3 mi Island nuclear plant in Pennsylvania destroyed public and political confidence in nuclear power, which had been seen as a technological solution to restoring the declining commodity cost and financial strength of the electricity enterprise. This event fell immediately on the heels of the nuclear accident-themed movie, *The China Syndrome,* and seemed to validate nuclear power plant risks in the public mind. Although the lack of any core meltdown or even radiation leakage was testament to the quality and integrity of nuclear power plant design and construction, the demand for stricter safety regulations led to rapid cost escalation.

The first commercial nuclear power unit built in 1957 had a rating of 60 MW. By 1966, utilities were ordering units larger than 1000 MW, even though manufacturers had no experience with units larger than 200 MW at the time. This arguably over-aggressive design-by-extrapolation, plus uneven utility operations and maintenance (O&M) training and management, led to reactor cost overruns that were sending power companies to the brink of bankruptcy while average power prices soared 60% between 1969 and 1984. Utilities and manufacturers in 1965 predicted that 1000 reactors would be operating by 2000 and providing electricity "too cheap to meter." The reality was that only 82 nuclear plants were operating in

Elec–Elec

2000, and no new U.S. orders had been placed in two decades.

These issues were profoundly impacting the electricity enterprise in the 1980s. The very ways electricity was generated and priced were being challenged for the first time in nearly a century. No longer could planners count on a steady rise in electricity demand. No longer could utilities count on low-cost fuels or the promise of the atom. They could no longer construct larger and more efficient generation, nor could they avoid the costs associated with environmental emissions. Competition from innovative technologies and hustling entrepreneurs could no longer be blocked, and the long-standing social contract consensus among the stakeholders of the electricity enterprise began to unravel.

The push for open power markets started when energy-intensive businesses began demanding the right to choose their suppliers in the face of rising electricity prices. Recognizing this pressure, the Energy Policy Act of 1992 also greased the skids for greater competition. The Act let new unregulated players enter the electricity generation market and opened up the use of utilities' transmission lines to support wholesale competition. The deregulation of natural gas in the 1980s made gas a more available, affordable, and cleaner fuel for electricity generation. This, coupled with rapid advancements in aircraft-derivative combustion turbines, provided an attractive vehicle for new, independent power producers to enter the market with low capital investment. Notably, non-utility sources as late as 1983 supplied only 3% of the U.S. generation market. By 2003, however, unregulated non-utility generators had captured nearly 30% of the U.S. generation market, exceeding the combined share from rural coops, the federal government, and municipal utilities. Wholesale electricity trading also soared—from approximately 100 million kWh in 1996 to 4.5 billion kWh in 2000.

Between 2000 and 2003, about 200,000 MW of new natural gas-fired combustion turbine capacity were added to the U.S. electricity generating fleet. Sixty-five percentage of this new deregulated generating capacity utilizes a combined-cycle technology. However, the rate of new combustion turbine-based capacity addition has dropped off dramatically since then. In addition, the performance of this new capacity has suffered in terms of both heat-rate and capacity factor. These are all symptoms of the boom-bust cycle in power generation that now exists in the restructured industry. For example, the average capacity factor for the fleet of new combined-cycle power plants dropped from 50% in 2001 to below 30% in 2004 as electricity supply capability significantly exceeded demand, market access was physically limited by transmission constraints, and natural gas prices rose dramatically.

The 21st century has not begun well for the performance and integrity of the U.S. electricity enterprise. The biggest power marketer, Enron, collapsed amid scandal while facing a slew of lawsuits. Pacific Gas and Electric, one of the largest investor-owned utilities, filed for bankruptcy amid the chaotic power markets in California. 50 million people in the Northeast and Midwest lost electricity because of a cascading power failure in 2003 that could have been prevented by better coordination among utility operators. High natural gas prices and the U.S. economy's overall slowdown caused electricity demand to falter and wholesale prices to fall. As noted, the natural gas-fired combustion turbine boom collapsed, carrying many independent generation companies with it.

The end result was a $238 billion loss in market valuation for the electricity enterprise by early 2003 and the worst credit environment in more than 70 years. Those companies able to maintain good credit ratings and stable stock prices bought nearly $100 billion in assets from weaker firms. Since the Energy Policy Act of 1992, competitive electricity generators have been able to charge market rates, while the transmission and distribution sides of the enterprise have remained regulated with relatively low investment returns. As a result, more power is being produced, but it is being sent over virtually a frozen grid system. The U.S. Department of Energy predicts that transmission investment is only likely to expand 6% compared to the 20% growth in electricity demand expected over the coming decade. Another rising controversy pits residential against business customers as electricity rates increase.

In summary, as shown in Fig. 3, the past 30 years have focused on efforts to restore the electricity enterprise's declining cost commodity tradition. All have failed to meet this challenge and there are no "silver bullets" on the horizon that are likely to change this reality within the context of today's aging electricity supply infrastructure. At the same time, electricity has become increasingly politicized as an essential retail entitlement where market price volatility is effectively allowed to

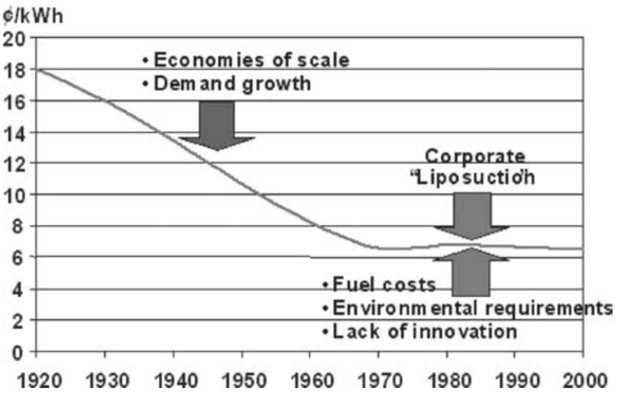


Fig. 3 Average U.S. price of residential electricity service (in 1984 $).

Elec–Elec

operate only in one direction—downward. Thus, the essential foundation for restoring vitality to the electricity enterprise rests first and foremost on innovation, principally in the consumer/delivery interface and in end-use electro-technologies. This represents a profound transformational challenge for the enterprise, which, throughout its history following the Edisonian beginning, has focused on supply-side technology as the wellspring of progress.

This combination of rising costs and artificially constrained price creates an economic vise on electricity supply that is squeezing out more and more value from the enterprise and the nation. Unfortunately, the dominant financial imperative has been to contain immediate costs at the expense of infrastructure development and investment. Unless the resulting standstill is ended and the assets of the enterprise are urgently reinvented, they risk being left behind as industrial relics of the 20th century. For example, the total capital expenditure rate of the electricity enterprise, both regulated and unregulated, as a fraction of its electricity revenues is now about 10%, less than one-half of the historic minimum levels and, in fact, a percentage only briefly approached during the depths of the Depression (refer to Fig. 4).

Unfortunately, this emphasis on controlling costs at all cost has also resulted in a period of profound technological stasis throughout the grid-based electricity enterprise. This has not been for lack of innovative opportunity but rather the lack of financial incentives. Every aspect of the enterprise has both the need and opportunity for technological renewal. For example, although coal remains the backbone of the power generation fleet producing over half the nation's electricity, the outdated technology being used is both inefficient and unable to keep pace with rising environmental demands, including carbon control. Integrated coal gasification-combined cycle (IGCC) technology which, in effect, refines coal into a clean synthesis gas for both electric power generation and synthetic petroleum production, could fundamentally address these constraints. Similarly, advanced nuclear power cycles could resolve the waste management, proliferation, and efficiency problems limiting the use of this essential clean energy source. Without considerably greater R&D emphasis and modernization of the power delivery system, renewable energy will also remain a theoretically attractive but commercially limited resource opportunity. The same is true of electrical storage.

Throughout the history of commercial electrification, large-scale storage, the long-sought after "philosopher's stone" of electricity, has remained elusive and thus a fundamental constraint on addressing optimal load management and asset utilization. Pumped storage, the use of off-peak electricity to pump water behind dedicated hydroelectric dams, has gained acceptance where feasible within geologic and environmental constraints. Demonstrations of compressed air storage using evacuated salt domes and aquifers were also successful, although this technology has not yet achieved significant commercial acceptance. At the other end of the spectrum of electricity storage, small-scale devices, including batteries and capacitors, are used for short-term load stability purposes and to dampen current fluctuations that affect reliability. A variety of advances in storage, including super conducting magnetic energy storage (SMES) and flywheels are being explored, but all have suffered from the general technological malaise constraining the grid-based electricity enterprise.

Only in applications for power portability has innovation in electricity storage made significant technical and commercial progress during this period. This progress also represents quite a different set of players than those of the traditional electric utility-based industry. Also indicative

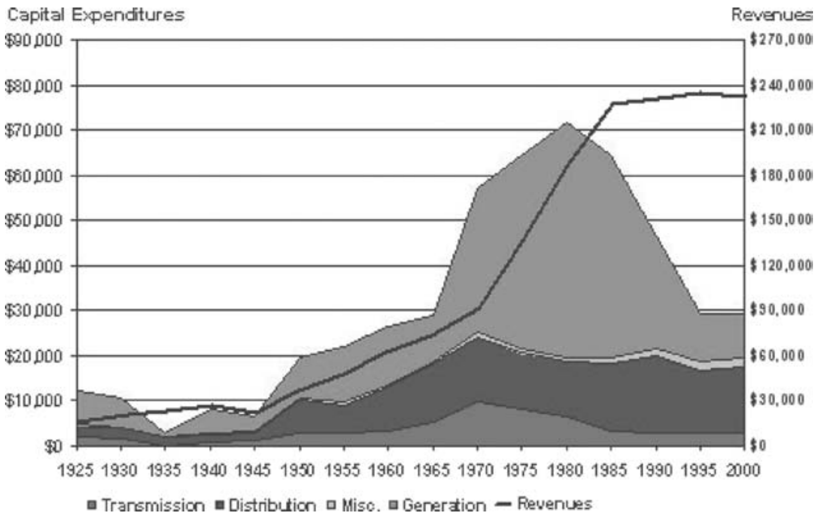


Fig. 4 Electric utility revenues and capital expenditures (in 2003 millions of $), 1925–1999.

of new players taking advantage of the growing value gaps in the nation's traditional electricity supply capability is the trend toward distributed power resources. Consumers with urgent needs for high quality power are increasingly taking advantage of the emergence of practical on-site generation technologies. These include diesel power sets, microturbines utilizing natural gas or landfill methane, fuel cells, and hybrid power systems incorporating photovoltaics. All of these are, in effect, competitors with the power grid, although ideally they could be integrated as grid assets within a truly modernized national electricity service system.

Similarly, in terms of power delivery technology, a wide array of thyristor-based digital control systems (e.g., FACTS), wide-area monitoring and communications, and highly sensitive anticipatory condition monitors, have been demonstrated and could revolutionize the reliability, capacity, and operability of the nation's electricity transmission and distribution network. Superconductivity represents another potential breakthrough technology that could fundamentally improve the efficiency of both power delivery and end use. This has been enabled by the recent development of so-called "high-temperature" superconductive materials operating at relatively modest liquid nitrogen temperatures. These materials, in effect, have no electrical resistance, thus eliminating transmission distance limitations, and are capable of significantly increasing the electrical capacity of existing distribution systems. The primary constraint today is the brittle ceramic nature of these superconducting materials and the resulting difficulties in manufacturing durable wiring, etc.

However, unless and until the electricity system advances from its current "life-support" level of infrastructure investment, all these potential advances remain, at best, on-the-shelf novelties. (These and others will be addressed further in the following entry on Transformation.) This investment gap is exacting a significant cost that is just the tip of the iceberg in terms of the electricity infrastructure's growing vulnerabilities to reliability, capacity, security, and service challenges. In fact, since the mid-1990s, the electric utility industry's annual depreciation expenses have exceeded construction expenditures. This is typical of an industry in a "harvest the assets" rather than an "invest in the future of the business" mode.

An even more dramatic measure of stasis is the minimal R&D investment by the electricity enterprise. In the wake of restructuring in the early 1990s, the enterprise's R&D investment rate has declined to about 0.2% of annual net revenues. This compares to a U.S. industry-wide average of about 4%. Even with inclusion of federal electricity-related R&D, the total is still only equivalent to a fraction of one percent of annual electricity sales revenues. The bill for this mortgaging of the future has come due and will, unless promptly paid, impose a heavy price on the nation's productivity, economy, and the welfare of its citizens.

Fig. 5 compares recent relative price trends among a variety of essential retail consumer goods and services. On the surface, the trend for electricity looks quite favorable but, after factoring in the rapidly growing cost of service unreliability (the shaded area), the real cost of electricity service has been increasing significantly and continues to do so at an escalating rate. This is an indelible reminder that it is only the lack of quality that adds to cost.

There is also a significant and growing stakeholder concern that the electricity enterprise, as currently constituted, is out of step with the nation at large. As economic growth resumes, will the enterprise be able to keep pace with energy quantity and quality needs, and can it also satisfy investor expectations without again resorting to questionable, high-risk financial schemes?

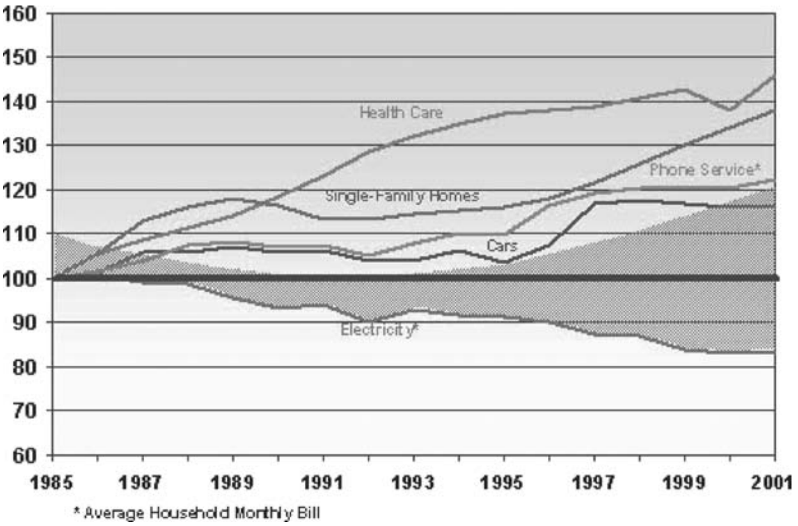


Fig. 5 Relative consumer prices, 1985–2001.

Table 4 Maturity and stasis

	1968	1985	2002
Ultimate customers (in millions)	70	101	132
Net generation (10^9 kWh)	1,436	2,545	3,841
Installed generating capacity (10^3 MW)	310	712	981
Average power plant size (MW)	85	227	300
Circuit miles of high-voltage line[a] (10^3 mi)	425	605	730
Residential service price (¢ per kWh—1983$)	7.1	6.8	6.8

[a] 22,000 V and above.

Table 4 summarizes the course of the enterprise during the maturity and stasis period of the last 35 years.

CONCLUSION

The past century has witnessed the creation of the electric utility industry, and the profound growth in electric use. Electricity is now consumed by virtually every residential, commercial, industrial, and institutional facility in the United States. For the first half century of this new technology, costs of electricity continually declined; and by the 50 year mark, a unified electric grid system reached the far corners of the nation. The electric power system in the U.S. had reached its zenith. But, by the mid-1970s, the electric industry was hit by several price shocks—the most damaging being the end of ever cheaper power plants and ever cheaper electric power. The more recent trend toward deregulation in the electric industry has had a significant impact on electricity customers, as well as on electric utilities. Cost increases for customers, economic problems for utilities, and reliability problems for the electric grid have all become serious problems for the electricity enterprise. A "reinvention" of the electricity enterprise is needed to control costs, increase the economic health of electric utilities, and prepare for the future uses of electricity. A modernized electricity enterprise would provide widespread benefits for the U.S. economy and society. The need for a transformed electricity enterprise in the U.S. is the topic of a second article in this area.

REFERENCE

1. *Statistical Yearbook of the Electric Utility Industry*; Edison Electric Institute: Washington, DC, 2003.

BIBLIOGRAPHY

1. Brennan, T.J. et al. *Alternating Currents—Electricity Markets and Public Policy*; Resources for the Future: Wahington, DC, 2002.
2. *Electricity Sector Framework for the Future*; Electric Power Research Institute (EPRI): Palo Alto, CA, 2003, www.epri.com/corporate/esff/default.asp.
3. *Electricity Technology Roadmap*; Electric Power Research Institute (EPRI): Palo Alto, CA, 2003.
4. Gellings, C.; Yeager, K. *Transforming the Electricity Infrastructure, Physics Today*; American Institute of Physics: College Park, MD, 2004; 45–51.
5. Hirsch, R.F. *Technology and Transformation in the American Electric Utility Industry*; Cambridge University press,: Cambridge, UK, 1989.
6. Hyman, L.S. *America's Electric Utilities—Past, Present and Future*, 6th Ed.; Public Utilities Reports, Inc.: Arlington, VA, 1997.
7. Nye, D.E. *Electrifying America*; The MIT Press: Cambridge, MA, 1990.
8. Patterson, W. *Transforming Electricity*; The Royal Institute of International Affairs/Earthbeam Publications Ltd: London, 1999.
9. *Reviving the Electricity Sector*; National Commission on Energy Policy: Washington, DC, 2003, www.energycommission.org.
10. Schurr, S.H. et al. *Electricity in the American Economy*. Contributions in Economies and Economic History-117, Greenwood Press: New York, 1990.
11. Smi, V. *Energy at the Crossroads*; The Royal Institute of International Affairs/Earthbeam Publications Ltd: London, 1999.
12. Yeager, K. *The Widening Technological Divide*; Public Utilities Fortnightly: Washington, DC, April 2005; 54.

Electricity Enterprise: U.S., Prospects

Kurt E. Yeager
Galvin Electricity Initiative, Palo Alto, California, U.S.A.

Abstract
This entry examines the present status of the U.S. electricity enterprise, and seeks to identify future opportunities for technological innovation in the electric power systems (as broadly defined) that will best serve the changing needs of consumers and businesses over at least the next 20 years. Of paramount importance will be insuring that the electricity system provides absolutely reliable and robust electric energy service in the context of ever-changing consumer needs.

INTRODUCTION

The stages in the Electricity Sector Life-Cycle shown in Fig. 1. The first three stages of the life-cycle are discussed in another entry. This entry discusses the fourth stage of Transformation, and suggests a path to the future of the U.S. Electricity Enterprise.

Electricity can be used in ways that no other energy form can be used. In addition to new products and new techniques of production, technological progress over the past century has led to radically improved ways of organizing productive activities as well, and all of these innovations have been heavily dependent on electricity. As a result, electricity has become the lifeblood of the nation's prosperity and quality of life. In fact, the U.S. National Academy of Engineers declared that "the vast networks of electrification are the greatest engineering achievement of the 20th Century."

Looking to the future, this entry discusses the performance changes within the U.S. electricity enterprise needed to respond to the rapidly growing reliability and service value expectations of 21st century consumers and society. One possible approach to such a transformation is described here as an example. Ultimately, a broad range of potential technological innovations bearing on the future of the electricity enterprise, from both the supply and utilization perspective, will need to be examined. Based on that broad objective examination, a comprehensive blueprint for elevating the reliability and value proposition of electric energy service will need to be formulated.

Above all, such a transformed system must be able to remain robust in the face of future complications of all sorts. It should therefore incorporate mechanisms for learning, innovation, and creative problem solving well beyond the capabilities of the current electric energy service system. The robustness of the system will also require that the initiative and its participants consider the role of the system's evolutionary history in determining its current and future state, and the set of strategic options open to the system (*Robust Design: A Repertoire of Biological, Ecological and Engineering Case Studies*, edited by Erica Jen, Santa Fe Institute Studies in the Science of Complexity, Oxford University Press (2005). This book uses robustness as follows: "Robustness is an approach to feature persistence in systems that compels us to focus on perturbations, and assemblages of perturbations, to the system that are different from those considered in its design, or from those encountered in its prior history.")

TRANSFORMATION

The first step in restoring the integrity and building the value of the electricity enterprise, in the context of 21st century needs, is to focus on the fundamentals that stakeholder input has helped to highlight:

1. Electricity is more than a form of commoditized energy; it is the underpinning of the modern quality of life, and the nation's indispensable engine of prosperity and growth.
2. Electricity is a service-based enterprise whose value to consumers depends on the most technically complex machines ever built.
3. The opportunities for technology to relieve cost pressures are principally in its ability to increase the service value of electricity. Building service value, over and above electricity's historic commodity energy value, is essential to every element of, and participant in, the electricity value chain.
4. The ability to capitalize on these new value opportunities requires a transformation of today's electricity infrastructure. This transformation must

Keywords: Electric enterprise; Restructuring; Electric system infrastructure; Grid system; Electric system transformation; Electric system stakeholders; Modern grid initiative; Galvin electricity initiative; Power quality; Digital quality power; Digital revolution; Portable power; Intelligrid.

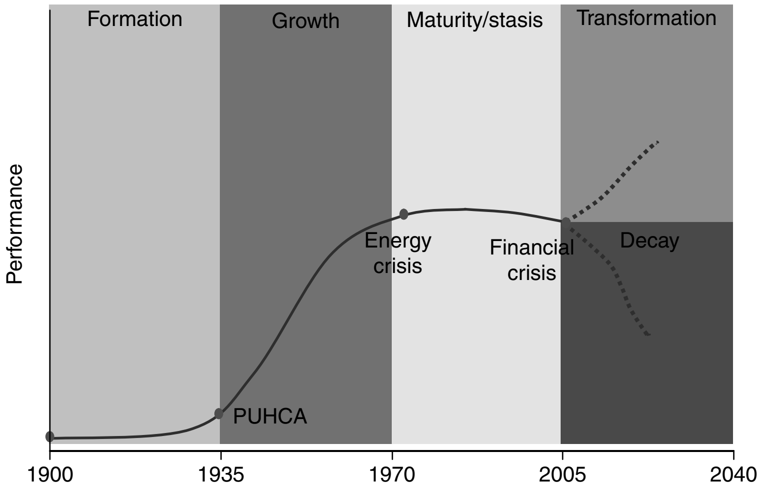


Fig. 1 Electricity sector life-cycle—a fork in the road.

enable all consumers to become active participants in, and benefactors of, the electricity enterprise, rather than remaining captive to the historic model of energy as strictly a commodity. Consumers want more choice and control.

The stakeholders' feedback also underscores that this "forward-to-fundamentals" transformation of the electricity enterprise is a process, not an event. This process should develop and proceed based on local benefits and costs—just as the nation's highway system was developed in the 20th century. It will also be a process because it requires a transformation of the institutional as well as physical infrastructure.

This "reinvention" of the electricity enterprise is likely to be as daunting as it is essential. Indeed, it is more typical for incumbents to consider innovative transformation as a greater threat to their status than as an opportunity. However, the current conditions prevailing in the enterprise and the growing service-value gap created by these conditions provide a situation in which transformative change is rapidly becoming a matter of survival for many incumbents, in addition to being a national imperative for productivity.

Unfortunately, there are very few examples of successful transformative reinvention of systems by established incumbents in any endeavor. The issue is even more challenging in the case of the electricity enterprise because of the significant barriers to entry by new players. Although the Public Utility Holding Company Act (PUHCA) was repealed in the just-passed Energy Policy Act of 2005, this is unlikely to attract significant new entrants. That is, unless the prevailing regulatory conflicts, uncertainties, and other disincentives are resolved that effectively create an impenetrable "iron curtain" around the enterprise, and transformational change leadership is mobilized.

Transformation requires that both the industry and its regulators move beyond their traditional cost-plus (or minus) commodity culture in terms of electricity's value proposition to consumers and society. A compelling, unified leadership vision is needed that conveys the fact that electricity, through innovative technology, provides a service value to consumers and society that is significantly greater than its basic commodity value. In order to realize this vision, regulatory policy must facilitate true consumer participation by urgently resolving the tension in the changing role of regulation from one primarily of protecting to one of enabling.

The foundation for confidence in this enterprise transformation stems from the revolution underway in the enabling technologies affecting all network infrastructures. A portfolio of innovative technologies can comprehensively resolve the vulnerability of today's power supply system in terms of its capacity, reliability, security, and consumer service value. These "smart technologies" will also open the door to fully integrating distributed resources and central station power into a single network, in a manner than can reduce system vulnerability rather than add to it—as is typically the case today—while also steadily improving the efficiency and environmental performance of the system.

Since 1980, the electricity intensity (kilowatt-hours per dollar GDP) of the U.S. economy has declined about 10%, leaving the intensity today about where it was three and a half decades ago. Is this the result of real efficiency improvement or just another measure of stasis? The electricity enterprise stands at a critical fork in its road of progress. Today its stakeholders have the necessity, the opportunity, and the means to make a clear choice about the future value of the enterprise. The decisions made and the path taken will make a profound difference, not only to the destiny of the electricity enterprise, but also to the nation, and ultimately to the world.

A primary rationale behind the restructuring of the electric utility industry ten years ago was that competitive markets manage supply and demand, provide incentives for innovation, and allocate investments more effectively than centrally regulated monopolies. While fundamentally sound in principle, the policy implementation of this rationale has not adequately reflected either the unique physics or the public entitlement characteristics of electricity. The consequence has been a breakdown in the traditional public/private partnership built around the obligation to reliably serve, and upon which the value and reputation of the electricity enterprise was built. The decision to begin the competitive market transformation of electricity with the wholesale supply function rather than with retail service has proven to be particularly counter-productive. As a result, most of the potential consumer benefits of innovation have been left on the table thus seriously compromising the consumer value of a transformation.

In short, the electricity enterprise has tended, through restructuring, to become a victim of its historic success in maintaining universal service reliability at ever-lower costs. The essential foundation for restoring enterprise vitality in the coming decade is rebuilding this fundamental public/private partnership, based on technology innovations that can increase the quality and value of electricity service, particularly providing higher levels of reliability and security. This transformation of the traditional electricity supply network into tailored, multi-functional service networks should also result in significant new business growth opportunities, reinforced by greater consumer satisfaction.

A modernized electricity enterprise would provide widespread benefits for the economy and society. In this enhanced scenario, productivity growth rates are higher and the economy expands more rapidly, while energy consumption, intensity, and carbon emissions are reduced relative to business as usual. Higher productivity rates can be sustained because a more reliable digital power delivery infrastructure would enable workers to perform existing and new functions more accurately and efficiently. This accelerated productivity growth has been demonstrated and established selectively in the economy, but its potential could be expanded dramatically through a smart power supply system. In effect, improved reliability and quality of electricity would enable the digital economy to expand at a broader and faster rate—an essential factor for successful U.S. competition in a global economy.

The Digital Revolution is the third major economic transformation in the last 100 years, and each has increasingly depended on electricity. Each also has created substantial new levels of wealth, as well as winners and losers at the scale of individuals, corporations, and nations—all depending on the effectiveness with which the innovative technology underpinning the economic transformation is exploited. In this new electricity business environment, it is the quality of customer connectivity and responsiveness that increasingly will differentiate the winners from the losers. The most precious business asset becomes the customer access portal. Every electricity supply function preceding it will be under relentless cost pressure, only moderated by the value that each consumer ultimately receives.

This transformed electricity infrastructure and business model will also serve to catalyze entirely new capabilities for consumer participation in the electricity marketplace, while significantly reducing the parasitic costs of power disturbances that are characteristic of today's system. This technological innovation will finally break open the commodity box currently constraining both the electricity enterprise and consumers, and will usher in an era of ever-higher valued energy/information services even beyond our imaginations. The payoff from this economic progress could easily exceed $1 trillion per year in additional U.S. GDP within a decade. This accelerated economic expansion is essential to meeting the nation's growing debt, security, and aging population costs.

Above all, this modernized electricity system would provide much greater functionality and service value for consumers. From a business perspective, this additional value is increasingly necessary to compensate for the expected significant rise in the cost of electricity. A number of upward trends are likely to increase electricity cost by at least 30% during the coming decade. These upward cost pressures largely occur because of forces beyond the control of the electricity enterprise, including fuel prices environmental protection regulation, and protection of the physical and cyber infrastructures against potential incursions.

The prices of fuels—natural gas, coal, and uranium—are rising, and are likely to remain significantly elevated. As a result, fuel costs are expected to account for at least half of the net increase in the price of electricity over this period.

Similarly, the cost of compliance with steadily tightening environmental regulations also continues to increase. In addition, there is a growing possibility that mandatory carbon control requirements will be instituted, possibly in the coming decade. This would have a significant additional impact on electricity costs, including the need for major strategic investments in cleaner replacement generation facilities using natural gas and nuclear, Integrated Gasification Combined Cycle (IGCC) for coal, and renewable resources.

Security improvements, both to discourage terrorist attacks and to recover from them, may further add to the cost of electricity. A fully functional security program would, for example, combine enhanced physical security systems with self-healing grid capabilities and cyber security advances. Cyber security would focus on increasing the security of communications, data monitoring, and automated system control functions, plus provide vulnerability assessments to support the self-correcting grid and related adaptive islanding capabilities. Rising fears of terrorist attacks may also fuel political pressure to further escalate infrastructure security measures, regardless of their cost. On the other hand, a modern, digitally monitored and controlled power supply system would comprehensively and confidently address these concerns as part of the process of technical modernization.

The need to improve the reliability and quality of power systems is another important reason the cost of electricity likely will increase through 2015. Mandatory reliability standards have been endorsed widely and were moving through Congress until stalled with the rest of the 2003 Energy Policy Act legislation. However, to be relevant to the needs of the new century, these standards will need to move beyond the traditional "keep the lights on" level of reliability to reflect the more stringent requirements of the digital economy.

Probably the greatest long-term challenge to the electricity sector is the fact that even as the demand for power is growing, the nature of electricity demand is undergoing a profound shift due to digital technology. Twenty years ago when the personal computer was introduced, few foresaw the widespread proliferation of "smart" devices. Today, for every microprocessor inside a computer, there are 30 more in stand-alone applications, resulting in the digitization of society. In applications ranging from industrial sensors to home appliances, microprocessors now number more than 12 billion in the United States alone.

These digital devices are highly sensitive to even the slightest disruption in power (an outage of less than a fraction of a single cycle can disrupt performance), as well as to variations in power quality due to transients, harmonics, and voltage surges and sags. "Digital quality power," with sufficient reliability and quality to serve these growing digital loads, now represents about 10% of the total electrical load in the United States, for example. It is expected to reach 30% by 2020 under business-as-usual conditions, and as much as 50% in a scenario where the power system is revitalized to provide universal digital-grade service.

However, the current electricity infrastructure in the United States, designed decades ago to serve analog (continuously varying) electric loads, is unable to consistently provide the level of digital quality power required by our digital manufacturing assembly lines, information systems, and soon even our home appliances. The economic loss of power disturbances mentioned earlier is attributable in part to the sensitivity of these new digital technologies.

Advanced technology now under development or on the drawing boards sustains the promise of fully meeting the electricity needs of a robust digital economy. The architecture for this new technology framework is becoming clear through early research on concepts and the necessary enabling platforms. In broad strokes, the architectural framework envisions an integrated, self-healing, electronically controlled electricity supply system of extreme resiliency and responsiveness—one that is fully capable of responding in real time to the billions of decisions made by consumers and their increasingly sophisticated microprocessor agents. In short, the potential exists to create an "IntelliGrid1" electricity supply system that provides the same reliability, efficiency, precision, and interconnectivity as the billions—ultimately trillions—of microprocessors that it will power. The following summarizes a potential set of key steps in this performance transformation:

- *Digitally controlling the power delivery network* by replacing today's relatively slow electro-mechanical switching with real-time, power-electronic controls. This will become the foundation of a new "smart, self-healing power delivery system" that will enable innovative productivity advances throughout the economy to flourish. Digital control is the essential step needed to most cost-effectively address the combined reliability, capacity, security, and market-service vulnerabilities of today's power delivery system. As a practical matter, this technical expansion is the only way that these vulnerabilities can be comprehensively resolved.
- *Integrating communications* to create a dynamic, interactive power system as a new "mega-infrastructure" for real-time information and power exchange. This is the capability needed to enable retail energy markets; power interactive, microprocessor-based service networks; and fundamentally raise the value proposition for electricity. Through advanced information technology, the system would be "self healing" in the sense that it is constantly self-monitoring and self-correcting to keep high-quality, reliable power flowing. It would sense disturbances and instantaneously counteract them, or reconfigure the flow of power to isolate any damage before it can propagate. To realize the vision of the smart power delivery system, standardized communications architecture must first be developed and overlaid on today's power delivery system. This "integrated energy and communications system architecture" should reflect an open, standards-based architecture for data communications and distributed computing infrastructure.
- *Automating the distribution system* to meet changing consumer needs. The value of electricity distribution system transformation—fully automated and integrated with communications—derives from four basic functionality advantages:
  - Reduced number and duration of consumer interruptions, system fault anticipation, and faster restoration.
  - Increased ability to deliver varying "octane" levels of reliable, digital-grade power.

— Increased functional value for all consumers in terms of metering, billing, energy management, demand-response, and security monitoring, among others.
— Access to selective consumer services including energy-smart appliances, power-market participation, security monitoring, and distributed generation.

To a power system operator, automation means a self-healing, self-optimizing smart power delivery system that automatically anticipates and quickly responds to disturbances, thus minimizing, if not ultimately eliminating, power disruptions altogether.

- *Transforming the meter* into a consumer gateway that allows price signals, decisions, communications, and network intelligence to flow back and forth through the two-way energy/information portal. This will be the linchpin technology that leads to a fully functioning marketplace with consumers responding (through microprocessor agents) to price signals. For consumers and providers alike, this gateway or portal through today's opaque electric service "iron curtain" provides the tool for moving beyond the commodity paradigm of 20th century electricity service. The result will quite possibly usher in a new set of energy/information services at least as diverse as those in today's telecommunications. This portal would sit between consumers' "in-building" communications network and wide-area "access" networks, enabling two-way, secure, and managed communications between consumers' equipment and energy service and/or communications entities.
- *Integrating distributed energy resources*. The new system would also be able to seamlessly integrate an array of locally installed, distributed power generation units (such as fuel cells and renewable resources) as power system assets addressing a variety of challenges. These challenges include the needs to increase the resiliency and reliability of the power delivery infrastructure, provide high-quality power, facilitate the provision of a range of services to consumers, and provide consumers with lower-cost, higher-quality power. These distributed power sources could be deployed on both the supply and consumer side of the energy/information portal as essential assets dispatching reliability, capacity, and efficiency. Unfortunately, today's electrical distribution system, architecture, and mechanical control limitations, in effect, prohibit this enhanced system functionality.
- *Accelerating end-use efficiency* through digital technology advances. The growing trend toward the digital control of processes can enable sustained improvements in efficiency and worker productivity for nearly all industrial and commercial operations. Similarly, the growth in end-use electrotechnologies, networked with system controls, will afford continuous improvements in user productivity and efficiency.
- *Expanding portable power*. The added value of individualized consumer electricity services is exemplified by the rapidly proliferating market for portable electricity devices and power supplies. These are likely to capture more and more of the higher-value electricity uses. More efficient power usage is reducing the overall power demands of these devices while new cell chemistries are expected to double usable power densities of these power supplies.

Lithium-ion battery technology is rapidly leading the demand for powering portable devices, while the market for nickel–cadmium (NiCad) is shrinking under environmental pressures. The price of lithium-ion batteries has also dropped by 20%–50% during the last few years, while NiCad and nickel-metal hydride battery prices have declined by 10%–20%. Primary batteries can be stored up to 10 years and have much higher energy densities than rechargeable secondary batteries.

Fuel cells are a particularly attractive alternative for greater power portability. However, until major cost, size, and performance breakthroughs are achieved, fuel cell use will remain limited in portable applications. Fortunately, this is an area of technology that is receiving substantial developmental investment to resolve these constraints.

With the switch to electronic automobile braking and steering by wire, the 3 kW capability of the rechargeable lead-acid and the single 12-V battery will no longer be sufficient, and is likely to usher in the 42-V system for automobiles. Hybrid vehicles require a high-voltage battery of about 150 V. This is currently provided by connecting nickel-metal-hydride cells in series. Battery life in this application is crucial since replacement costs as much as a new motor.

ACHIEVING THE TRANSFORMATION

There is growing recognition by the electricity sector of the need to modernize its aging infrastructure and business model. This is reflected in a series of complementary initiatives being pursued by various public and private organizations. These initiatives include the Modern Grid Initiative sponsored by the U.S. Department of Energy and its National Energy Technology Laboratory; the Gridwise Architectural Council sponsored by the Battelle Pacific Northwest National Laboratory; and the IntelliGrid Consortium sponsored by the Electric Power Research Institute. Each of these initiatives is intended to comprehensively modernize the nation's electric energy supply system within the current regulated utility structure.

All of these initiatives have been developed over the last several years in response to the concerns of the broad electricity stakeholder community about the nation's aging electricity supply system. This reflects the fact that the nation and its economy are dependent on the integrity of a complex web of digital networks for which the electric power system is the foundation. These initiatives reflect the broad-based collaboration of leaders in energy, technology, and government, working to address the looming electricity supply industry issues and set the United States on a migration path toward the intelligent, self-healing power system of the future.

The foundation of this new system is an open-systems-based, comprehensive reference architecture. This architecture enables the sustainable integration of intelligent equipment and data communications networks on an industry-wide basis. This intelligent system will enable real-time energy information and power exchanges, in addition to fundamentally enhanced system reliability and security. These initiatives apply the latest system engineering methods and computational tools to model the advanced energy enterprise of the future. In so doing, they cut across traditional operating boundaries, promoting greater interoperability, and enabling unprecedented improvement in performance and asset utilization.

The result will be an electricity and information mega-infrastructure that enables technological advancement to continue to flourish. This system will be "always on and active," and interconnected into a national network of real-time information and power exchanges. This system will also have the intelligence to seamlessly integrate traditional central power generation with distributed energy resources.

The foundational reference architecture of this transformed electricity supply system also eliminates the constraints on the consumer gateway to the system now imposed by the meter. The result will allow price signals, communications, and network intelligence to flow back and forth through a two-way portal. This is the linchpin enabling a fully functioning retail power marketplace with consumers responding (through microprocessor agents) to price signals. Specific capabilities can include the following:

- Pricing and billing processes that support real-time pricing.
- Value added services such as billing inquiries, service calls, emergency services, and diagnostics.
- Improved building and appliance standards.
- Consumer energy management through sophisticated on-site systems.
- Easy "plug and play" connection of distributed energy resources.
- Improved real-time system operations.
- Improved short-term local forecasting.
- Improved long-term planning.

These grid modernization initiatives are being designed to accommodate widespread deployment of distributed energy resources (DER). DER includes microturbines, fuel cells, photovoltaics, and energy storage devices installed close to the point of use—all of which can reduce the need for intensive power delivery system investment. Particularly suitable for a variety of industry and commercial applications, DER reinforces the bulk power system by providing consumers with lower cost peak power, higher reliability, and improved power quality. Many DER devices also provide recoverable heat suitable for cogeneration.

In summary, these various grid modernization initiatives are intended to ultimately provide the technological and engineering means to resolve the large and growing gap between the performance capability of today's bulk electric energy supply system and the needs and expectations of modern society. Also fundamental to resolving this gap is a corresponding commitment to restore the level of infrastructure investment needed to fully develop and deploy the essential technological advancements.

Recognizing the deeply entrenched resistance to transformative change that exists in established enterprises—particularly such as the regulated monopoly structure of the electricity sector, another complementary initiative is taking a fundamentally different approach to achieving electricity system transformation.

In contrast to the various grid modernization initiatives described above, the Galvin Electricity Initiative seeks to transform the electric energy supply system starting from the consumer rather than from the supplier. This recognizes that the most important asset in resolving the growing electricity cost/quality dilemma and its negative reliability, productivity, and value implications is consumer-focused innovation that disrupts the performance status-quo by targeting and quickly demonstrating the advantages of system modernization.

The goal of the Galvin Electricity Initiative is the "Perfect Power System". That is, a system that never fails to meet, under all conditions, each consumer's expectations for electricity service confidence, convenience, and choice. In the context of this Initiative, the electric energy supply system includes all elements in the chain of technologies and processes for electricity production, delivery, and use across the spectrum of industrial, commercial, and residential applications. This focus on the consumer also reflects the relatively intractable nature of the highly regulated monopoly of bulk power infrastructure that dominates U.S. electric energy supply and service.

In addition to absolute quality, a second principle guiding the Galvin Electricity Initiative is enabling self-organizing entrepreneurs to engage in the nation's electricity enterprise. Innovative entrepreneurial leadership guided by consumer service opportunities is seen by the Initiative as providing the most confident engine for quality transformation and sustainable system improvement. The emphasis here is on creative, "outside the box" thinking that focuses on achieving maximum consumer value through innovation with as short a turnaround as possible.

The basic approach to developing the Perfect Power System is to increase the independence, flexibility, and intelligence of electric energy management and use at the device level and then integrate these capabilities at larger scales as necessary to achieve the delivery of perfect power. In this context, the Initiative is developing three levels of generic electric energy system configurations that have the potential to achieve early perfection, together with the corresponding technology innovation opportunities that are essential to their success. These configurations are: (a) Device-Level (portable) systems serving a highly mobile digital society; (b) Building Integrated Systems which focus on modular facilities serving individual consumer premises; and (c) Distributed Microgrid Systems including interconnection with local bulk power distribution networks. It is at this largest configuration level that the Galvin Electricity Initiative seamlessly interfaces with the modernization initiatives of the bulk power system described previously. Thus, each level of Galvin Perfect Power configuration reflects a consumer-guided step on the path to ultimately transforming the nation's entire bulk electric energy supply system.

In the context of each system configuration level, the Galvin Electricity Initiative is also performing comprehensive evaluations of eight target innovation nodes and their enabling technologies. The innovation nodes being considered are: communications; computational ability; distributed generation; power electronics and controls; storage; building systems; efficient appliances and devices; and sensors. These evaluations include, in each case, quantifying the performance gaps that must be filled to achieve perfection, and defining the opportunities and risks involved in resolving these gaps within a 10–20 year period. Business opportunity templates plus a quality leadership/management guidebook and associated training courses are also being developed. In total, the results will provide a comprehensive and confident roadmap for achieving prompt, successful commercial implementation of this consumer-focused electric energy supply and service modernization Initiative.

Modern society is increasingly dependent on electric energy, and it expects individualized service that enables the intelligent control of energy consumption with a premium on reliability, efficiency, and environmental performance. Transformation to an interactive, consumer-directed, service capability will change the business dynamics of the electricity enterprise from one of simply providing electric energy as a bulk commodity to offering a portfolio of individualized energy services of far greater value. This consumer-focused quality transformation will also serve to restore innovation and investment, and resolve the growing vulnerabilities facing the regulated bulk electric energy infrastructure. Prototype commercial Perfect Power System installations using state-of-the-art enabling technologies are also being implemented to immediately demonstrate these performance advantages.

CONCLUSIONS

The overarching priority of the electricity enterprise and all its stakeholders is to pursue the policies and actions needed to stimulate system modernization. The current rate of investment in the power delivery system alone, some \$18–\$20 billion per year, is barely enough to replace failed equipment. To correct deficiencies in the existing system and to enable the smart power delivery system of the future will require double this amount annually. The result would be an electricity supply system fully capable of meeting the escalating needs and aspirations of the 21st century society. In summary, these needs and aspirations dictate a future where:

- The electricity enterprise confidently provides the nation's most essential platform for technical innovation, productivity growth, and continued economic prosperity. Eventually, it is expected that this platform will enable every end-use electrical process and device to be linked, in real-time, with the open marketplace for goods and services, including, but not limited to, electric power.
- Economic productivity increases substantially as a result of the transformation of the electricity enterprise, generating additional wealth to deal with the large societal, security, and environmental challenges of the 21st century.
- The roles, responsibilities, and rules governing the electricity enterprise have been clarified, enabling a revitalized public/private partnership that maintains confidence and stability in electricity sector financing. As a result, the rate of investment in the essential electricity infrastructure is substantially increased.
- The role of regulation has evolved from oversight of company operations and "protection" of ratepayers to oversight of markets, as well as enabling and guiding market transparency and specific public-good services (i.e., reliability standards, provider-of-last-resort, market transformation, etc.).
- National security and energy policies emphasize U.S. fuel diversity, placing electricity at the center of a strategic thrust to: (1) create a clean, robust

portfolio of domestic energy options including fossil, nuclear, and renewable energy sources, along with enhanced end-use efficiency; (2) develop a sustainable electric energy system providing the highest value to all consumers with perfect reliability; and (3) electrify transportation to reduce dependence on foreign oil.

The cost/benefits of electricity system modernization will be profoundly positive. For example, the cost to the average household would be less than $5 per month before taking any credit for the considerable energy-savings opportunities that each consumer would be empowered to achieve through a modernized, more functional power system. In return, as power reliability and quality improve, each consumer would save hundreds of dollars per year in the price of purchased goods and services. Even more financially significant would be the income potential added to each household as the nation's productivity, security, and competitiveness increase. Thus, the service quality improvements from electricity system modernization would not result in higher costs, but are the genesis of cost savings for all consumers. Again, improving quality always costs less.

In support of this more positive future, a proactive technology development program will be necessary. This should emphasize power system reliability and functionality, management of greenhouse gases, and the development of higher-value, more efficient, smart electricity end-use devices and services. Technological innovation will remain, as it has been throughout the history of commercial electricity enterprise, the essential asset determining the destiny of the electricity sector and its value to society.

The result is likely to be a profoundly transformed, multi-dimensional electricity service capability incorporating an array of distributed, stored, and portable power resources as assets. At the same time, this infrastructure transformation will enable the convergence of electricity, telecommunications, and sensors into a smart, sustainably robust, mega-infrastructure powering a universal digital society with absolute reliability. Most importantly, this transformation will enable an array of innovations in electricity service that is only limited by our imaginations.

In closing, the Galvin Electricity Initiative seeks to urgently catalyze this performance and value transformation. It will do so through a two-phase effort that first explores and evaluates the most promising opportunities for technological innovation throughout the electricity value chain, as seen from the broad consumer perspective. The goal here is a demonstrable and compelling perfection in electricity service quality that will mobilize the broad community of stakeholders to demand the necessary performance and value transformation. In its second phase, the Initiative will develop the comprehensive change in leadership plan for most effectively stimulating this electricity service and supply transformation. This plan will broadly consider and specifically recommend the essential policies, institutions, standards, and incentives, etc. needed to break today's pervasive innovation/investment logjam in the electricity enterprise. The results will truly electrify the nation and the world in the full meaning of the word.

A message of enduring vision:

So long as there remains a single task being done by men or women which electricity could do as well, so long will that development (of electrification) be incomplete. What this development will mean in comfort, leisure, and in the opportunity for the larger life of the spirit, we have only begun to realize.

—Thomas Edison, 1928

BIBILIOGRAPHY

1. Brennan, T.J.; Palmer, K.L.; Martinez, S. Alternating Currents—Electricity Markets and Public Policy, Resources for the Future, Washington, DC, 2002.
2. Electricity Sector Framework for the Future, Electric Power Research Institute (EPRI), Palo Alto, CA, www.epri.com/corporate/esff/default.asp; 2003.
3. *Electricity Technology Roadmap*, Electric Power Research Institute (EPRI), Palo Alto, CA , 2003.
4. Gellings, C.; Yeager, K. *Transforming the Electricity Infrastructure, Physics Today*; American Institute of Physics: College Park, MD, 2004; PP. 45–51.
5. Hirsch, R.F. *Technology and Transformation in the American Electric Utility Industry*; Cambridge University Press: Cambridge, U.K., 1989.
6. Hyman, L.S. *America's Electric Utilities—Past, Present and Future*, 6th Ed.; Public Utilities Reports, Inc.: Arlington, VA, 1997.
7. Nye, D.E. *Electrifying America*; The MIT Press: Cambridge, MA, 1990.
8. Patterson, W. *Transforming Electricity,* The Royal Institute of International Affairs; Earthbeam Publications Ltd.: London, 1999.
9. Reviving the Electricity Sector, National Commission on Energy Policy, Washington, DC, www.energycommission.org; 2003.
10. Schurr, S.H.; Burwell, C.C.; Devine, W.D., Jr.; Sonenblum, S. *Electricity in the American Economy, Contributions in Economies and Economic History-117*; Greenwood Press: New York, 1990.
11. Smil, V. *Energy at the Crossroads*; The Royal Institute of International Affairs, Earthbeam Publications Ltd.: London, 1999.
12. *Statistical Yearbook of the Electric Utility Industry*, Edison Electric Institute, Washington, DC, 2003.
13. Yeager, K. The Widening Technological Divide, *Public Utilities Fortnightly*; Washington, DC, 2005; 54 ff.

Electronic Control Systems: Basic

Eric Peterschmidt
Honeywell Building Solutions, Honeywell, Golden Valley, Minnesota, U.S.A.

Michael Taylor
Honeywell Building Solutions, Honeywell, St. Louis Park, Minnesota, U.S.A.

Abstract
Every piece of energy-consuming equipment has some form of control system associated with it. This article provides information about electronic control systems primarily used to control HVAC equipment. The same principles are used to control other equipment, such as lighting, compressed air systems, process equipment, and production equipment.

INTRODUCTION

Every piece of energy-consuming equipment has some form of control system associated with it. The controls can be as simple as a snap switch or as complicated as a dedicated microcomputer chip system. Larger pieces of equipment, along with buildings and industrial processes, typically use complex computer-based control systems to optimally control and operate them. This article provides information about electronic control systems primarily used to control HVAC equipment. However, the same technologies and principles are used to control other equipment, such as lighting, compressed air systems, process equipment and production equipment.

An electronic control system comprises a sensor, controller and final control element. The sensors used in electronic control systems are simple, low-mass devices that provide stable, wide-range, linear and fast response. The electronic controller is a solid-state device that provides control over a discrete portion of the sensor range and generates an amplified correction signal to control the final control element.

Features of electronic control systems include the following:

- Controllers can be remotely located from sensors and actuators.
- Controllers can accept a variety of inputs.
- Remote adjustments for multiple controls can be located together, even though sensors and actuators are not.
- Electronic control systems can accommodate complex control and override schemes.
- Universal type outputs can interface to many different actuators.
- Display meters can indicate input or output values.

The sensors and output devices (e.g., actuators, relays) used for electronic control systems are usually the same ones used on microprocessor-based systems. The distinction between electronic control systems and microprocessor-based systems is in the handling of the input signals. In an electronic control system, the analog sensor signal is amplified, then compared to a setpoint or override signal through voltage or current comparison and control circuits. In a microprocessor-based system, the sensor input is converted to a digital form, in which discrete instructions (algorithms) perform the process of comparison and control.

Fig. 1 shows a simple electronic control system with a controller that regulates supply water temperature by mixing return water with water from the boiler. The main temperature sensor is located in the hot water supply from the valve. To increase efficiency and energy savings, the controller resets the supply water temperature setpoint as a function of the OA (outdoor air) temperature. The controller analyzes the sensor data and sends a signal to the valve actuator to regulate the mixture of hot water to the unit heaters. These components are described in the section titled "Components."

A glossary of control system terms is given in the last section of this article.

Electronic control systems usually have the following characteristics:

Controller. Low voltage, solid state.
Inputs. 0–1 V d.c., 0–10 V d.c., 4–20 mA, resistance element, thermistor, thermocouple.
Outputs. 2–10 V d.c. or 4–20 mA device.
Control Mode. Two-position, proportional, proportional plus integral (PI) or step.

Keywords: Electronic control; HVAC controls; Automatic temperature controls; Sensors; Actuators; Controllers; Output devices; Indicating devices; Energy efficiency.

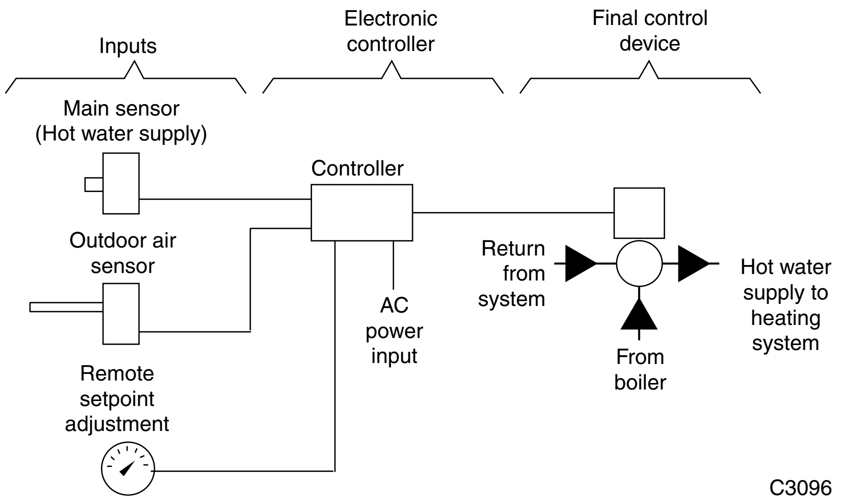


Fig. 1 Basic electronic control system.

Circuit diagrams in this article are basic and fairly general. A resistance-temperature input and a 2–10 V d.c. output are used for purposes of discussion. A detailed discussion on control modes can be found in the "Control Fundamentals" section of the Engineering Manual of Automatic Controls.[1]

ELECTRONIC CONTROL SYSTEM COMPONENTS

An electronic control system includes sensors, controllers, output devices such as actuators and relays; final control elements such as valves and dampers; and indicating, interfacing, and accessory devices. Fig. 2 provides a system overview for many electronic system components.

Sensors

A sensing element provides a controller with information concerning changing conditions. Analog sensors are used to monitor continuously changing conditions such as temperature or pressure. The analog sensor provides the controller with a varying signal such as 0–10 V. A digital (two-position) sensor is used if the conditions represent a fixed state such as a pump that is on or off. The digital

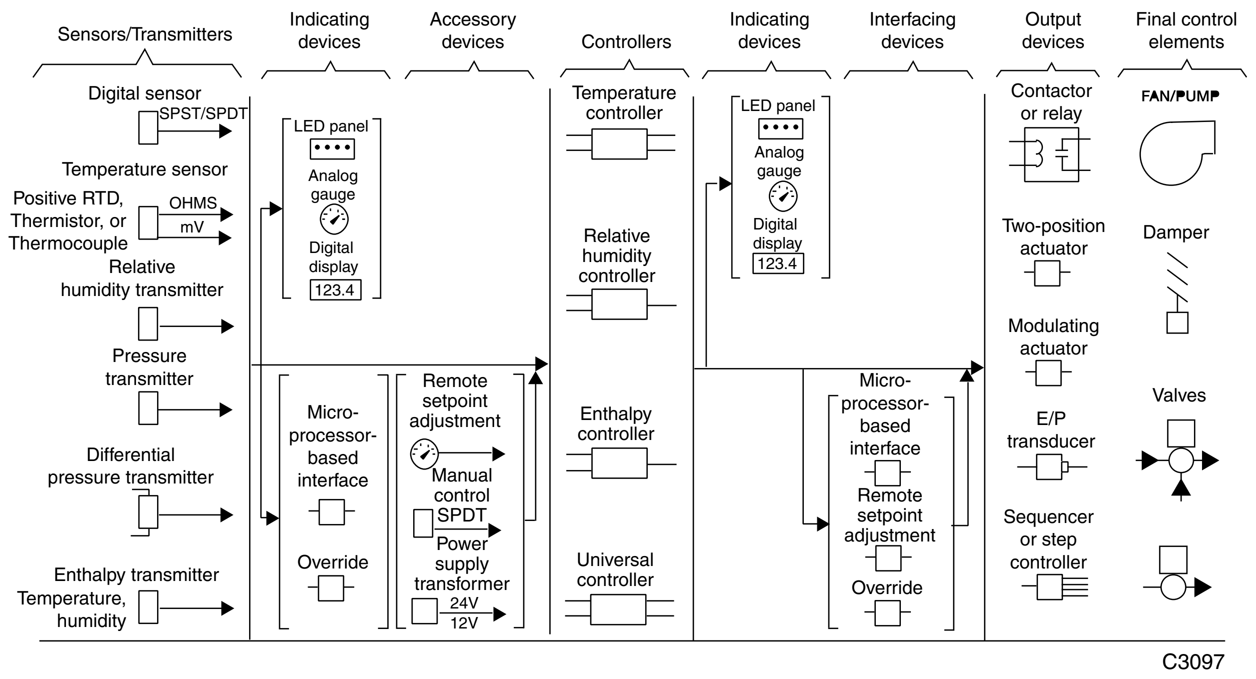


Fig. 2 Typical electronic control system components.

sensor provides the controller with a discrete signal such as open or closed contacts.

Some electronic sensors use an inherent attribute of their material (e.g., wire resistance) to provide a signal and can be directly connected to the electronic controller. Other sensors require conversion of the sensor signal to a type or level that can be used by the electronic controller. For example, a sensor that detects pressure requires a transducer or transmitter to convert the pressure signal to a voltage that can be used by the electronic controller. Typical sensors used in electronic control systems are included in Fig. 2. A sensor-transducer assembly is called a transmitter.

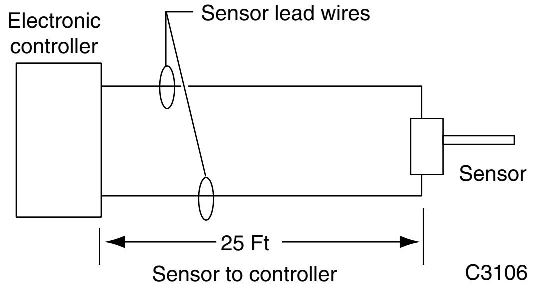


Fig. 3 Lead wire length.

Temperature Sensors

For electronic control, temperature sensors are classified as follows:

- Resistance temperature devices (RTDs) change resistance with varying temperature. RTDs have a positive temperature coefficient (resistance increases with temperature).
- Thermistors are solid-state resistance-temperature sensors with a negative temperature coefficient.
- Thermocouples directly generate a voltage as a function of temperature.

Resistance Temperature Devices

In general, all RTDs have some common attributes and limitations:

- The resistance of RTD elements varies as a function of temperature. Some elements exhibit large resistance changes, linear changes, or both over wide temperature ranges.
- The controller must provide some power to the sensor and measure the varying voltage across the element to determine the resistance of the sensor. This action can cause the element to heat slightly—called self-heating—and can create an inaccuracy in the temperature measurement. By reducing the supply current or by using elements with higher nominal resistances, the self-heating effect can be minimized.
- Some RTD element resistances are as low as 100 Ω. In these cases, the resistance of the lead wires connecting the RTD to the controller may add significantly to the total resistance of the connected RTD, and can create an offset error in the measurement of the temperature. Fig. 3 shows a sensor and controller in relation to wire lead lengths. In this figure, a sensor 25 ft from the controller requires 50 ft of wire. If 18 AWG solid copper wire with a d.c. resistance of 6.39 Ω/Mft is used, the 50 ft of wire has a total d.c. resistance of 0.319 Ω. If the sensor is a 100-ohm platinum sensor with a temperature coefficient of 0.69 Ω/°F, the 50 ft of wire will introduce an error of 0.46°F. If the sensor is a 3000-ohm platinum sensor with a temperature coefficient of 4.8 Ω/°F, the 50 ft of wire will introduce an error of 0.066°F.

Significant errors can be removed by adjusting a calibration setting on the controller, or—if the controller is designed for it—a third wire can be run to the sensor and connected to a special compensating circuit designed to remove the lead length effect on the measurement. In early electronic controllers, this three-wire circuit was connected to a Wheatstone Bridge configured for lead wire compensation. In digital controllers, lead wire compensation on low resistance sensors may be handled by software offset.

- The usable temperature range for a given RTD sensor may be limited by non-linearity at very high or low temperatures.
- RTD elements that provide large resistance changes per degree of temperature reduce the sensitivity and complexity of any electronic input circuit. (Linearity may be a concern, however.)

A sensor constructed using a BALCO wire is a commonly used RTD sensor. BALCO is an annealed resistance alloy with a nominal composition of 70 percent nickel and 30 percent iron. A BALCO 500-ohm resistance element provides a relatively linear resistance variation from −40 to 250°F. The sensor is a low-mass device and responds quickly to changes in temperature.

Another material used in RTD sensors is platinum. It is linear in response and stable over time. In some applications, a short length of wire is used to provide a nominal resistance of 100 Ω. However, with a low resistance value, the element can be affected by self-heating and sensor-leadwire resistance. Additionally, due to the small amount of resistance change of the element, additional amplification must be used to increase the signal level.

To use the desirable characteristics of platinum and minimize any offset, one manufacturing technique deposits a film of platinum in a ladder pattern on an

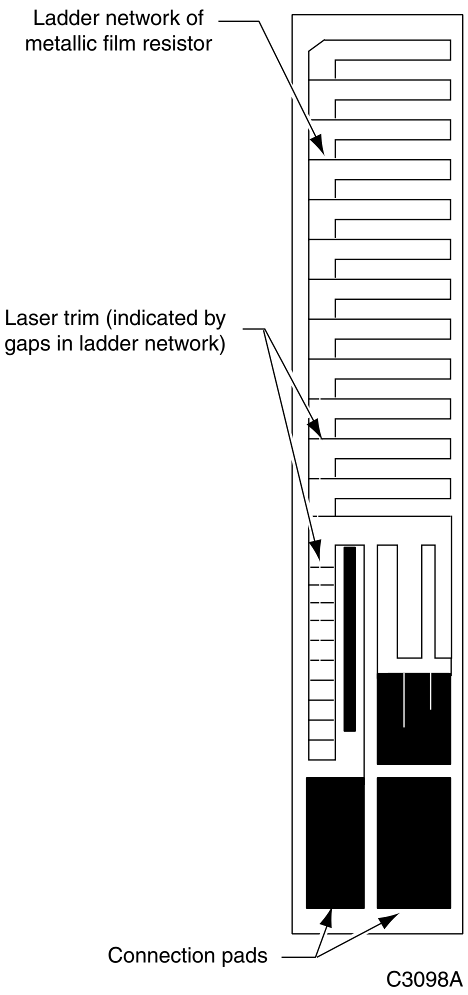


Fig. 4 Platinum element RTD sensor.

insulating base. A laser trimming method (Fig. 4) then burns away a portion of the metal to calibrate the sensor, providing a resistance of 1000 Ω at 74°F. This platinum film sensor provides a high resistance-to-temperature relationship. With its high resistance, the sensor is relatively immune to self-heating and sensor-leadwire resistance offsets. In addition, the sensor is an extremely low-mass device and responds quickly to changes in temperature. RTD elements of this type are common.

Solid-State Resistance Temperature Devices

Fig. 5 shows examples of solid-state resistance temperature sensors having negative and positive temperature coefficients. Thermistors are negative temperature coefficient sensors typically enclosed in very small cases (similar to glass diodes or small transistors) that provide quick response. As the temperature increases, the resistance of a thermistor decreases (Fig. 6). Selection of a thermistor sensor must consider the highly nonlinear temperature-resistance characteristic.

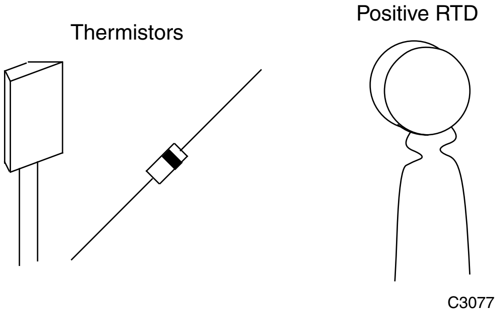


Fig. 5 Solid-state temperature sensors.

Positive temperature coefficient solid-state temperature sensors may have relatively high resistance values at room temperature. As the temperature increases, the resistance of the sensor increases (Fig. 6). Some solid-state sensors

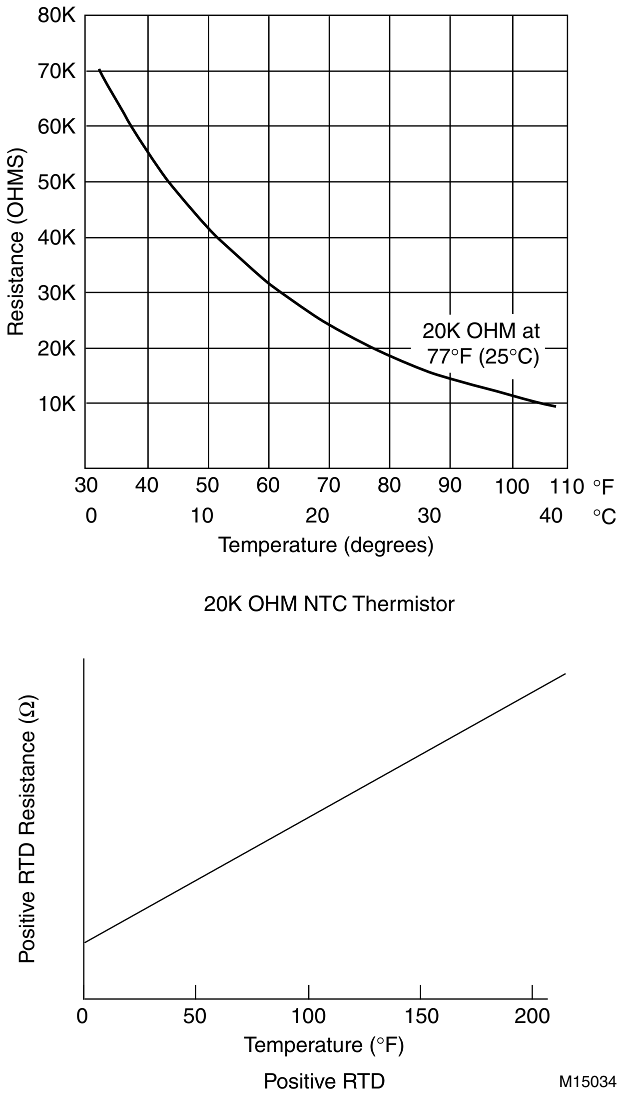


Fig. 6 Resistance vs temperature relationship for solid-state sensors.

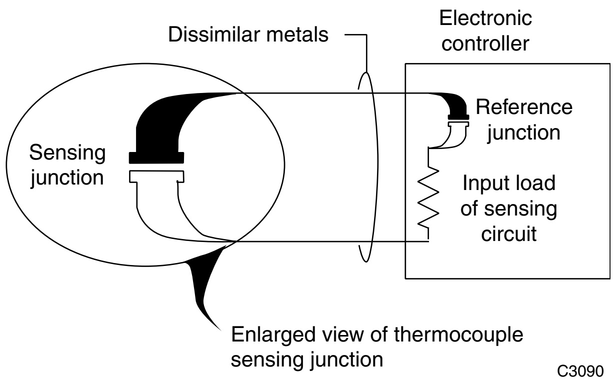


Fig. 7 Basic thermocouple circuit.

have near-perfect linear characteristics over their usable temperature range.

Thermocouples

In a thermocouple, two dissimilar metals such as iron and constantan are welded together to form a thermocouple junction (Fig. 7). When this junction is exposed to heat, a voltage in the millivolt range is generated and can be measured by the input circuits of an electronic controller. The amount of voltage generated is directly proportional to the temperature (Fig. 8). At room temperatures for typical HVAC applications, these voltage levels are often too small to be used, but are more usable at higher temperatures of 200°F–1600°F. Consequently, thermocouples are most common in high-temperature process applications.

Transmitter/Transducer

The input circuits for many electronic controllers can accept a voltage range of 0–10 V d.c. or a current range of

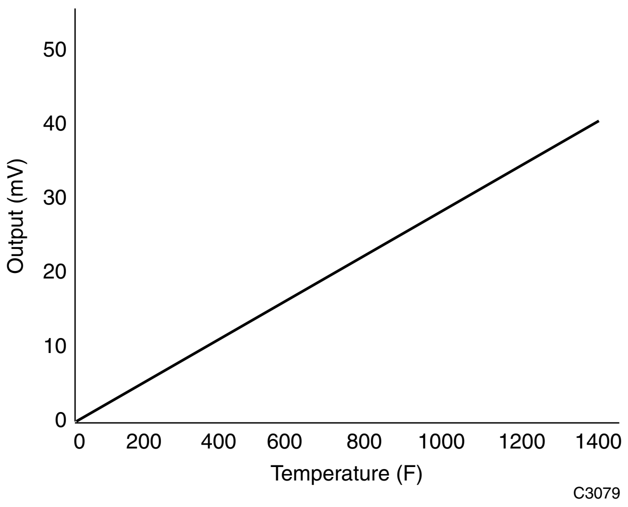


Fig. 8 Voltage vs temperature for iron-constantan thermocouple.

4–20 mA. The inputs to these controllers are classified as universal inputs because they accept any sensor having the correct output. These sensors are often referred to as transmitters as their outputs are an amplified or conditioned signal. The primary requirement of these transmitters is that they produce the required voltage or current level for an input to a controller over the desired sensing range.

Transmitters measure various conditions such as temperature, relative humidity, airflow, water flow, power consumption, air velocity and light intensity. An example of a transmitter would be a sensor that measures the level of carbon dioxide (CO_2) in the return air of an air handling unit. The sensor provides a 4–20 mA signal to a controller input, which can then modulate outdoor/exhaust dampers to maintain acceptable air quality levels. Since electronic controllers are capable of handling voltage, amperage or resistance inputs, temperature transmitters are not usually used as controller inputs within the ranges of HVAC systems due to their high cost.

Relative Humidity Sensor

Various sensing methods are used to determine the percentage of relative humidity, including the measurement of changes of resistance, capacitance, impedance and frequency.

An older method that used resistance to determine relative humidity depended on a layer of hygroscopic salt, such as lithium chloride or carbon powder, deposited between two electrodes (Fig. 9). Both materials absorb and release moisture as a function of the relative humidity, causing a change in resistance of the sensor. An electronic controller connected to this sensor detects the changes in resistance, which it can use to provide control of relative humidity.

A method that uses changes in capacitance to determine relative humidity measures the capacitance between two conductive plates separated by a moisture-sensitive material such as polymer plastic (Fig. 10A). As the material absorbs water, the capacitance between the plates decreases, and the change can be detected by an electronic circuit. To overcome any hindrance of the material's ability to absorb and release moisture, the two plates and their electric leadwires can be on one side of the polymer plastic, with a third sheet of extremely thin conductive material on the other side of the polymer plastic forming the capacitor (Fig. 10B). This third plate, too thin for attachment of leadwires, allows moisture to penetrate and be absorbed by the polymer, thus increasing sensitivity and response.

A relative humidity sensor that generates changes in both resistance and capacitance to measure moisture level is constructed by anodizing an aluminum strip and then applying a thin layer of gold or aluminum (Fig. 11). The anodized aluminum has a layer of porous oxide on its surface. Moisture can penetrate through the gold layer and fill the pores of the oxide coating, causing changes in both

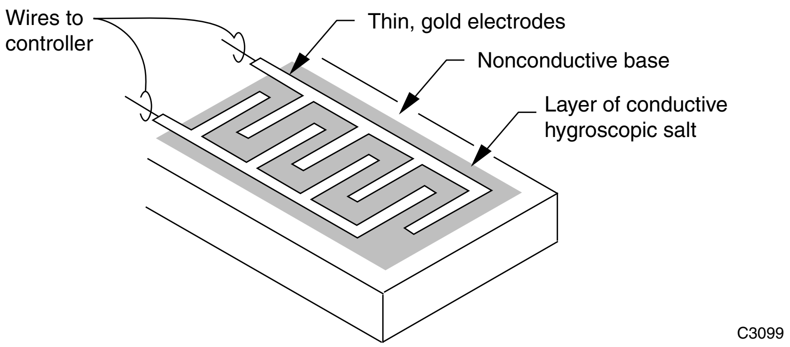


Fig. 9 Resistive-type relative humidity sensor.

resistance and capacitance that can be measured by an electronic circuit.

Sensors that use changes in frequency to measure relative humidity (Fig. 12) can use a quartz crystal coated with a hygroscopic material such as polymer plastic. When the quartz crystal is energized by an oscillating circuit, it generates a constant frequency. As the polymer material absorbs moisture and changes the mass of the quartz crystal, the frequency of oscillation varies and can be measured by an electronic circuit.

Most relative humidity sensors require electronics at the sensor to modify and amplify the weak signal and are referred to as transmitters. The electronic circuit compensates for the effects of temperature and both amplifies and linearizes the measured level of relative humidity. The transmitters typically provide a voltage or current output that can be used as an input to the electronic controller.

Pressure Sensors

An electronic pressure sensor converts pressure changes into a signal such as voltage, current or resistance that can be used by an electronic controller.

A method that measures pressure by detecting changes in resistance uses a small, flexible diaphragm and a strain gage assembly (Fig. 13). The strain gage assembly includes very fine (serpentine) wire or a thin metallic film deposited on a nonconductive base. The strain gage assembly is stretched or compressed as the diaphragm flexes with pressure variations. The stretching or compressing of the strain gage (shown by a dotted line in Fig. 13) changes the length of its fine wire or thin film metal, which changes the total resistance. The resistance can then be detected and amplified. These changes in resistance are small. Therefore, an amplifier is provided in the sensor assembly to amplify and condition the signal so the level

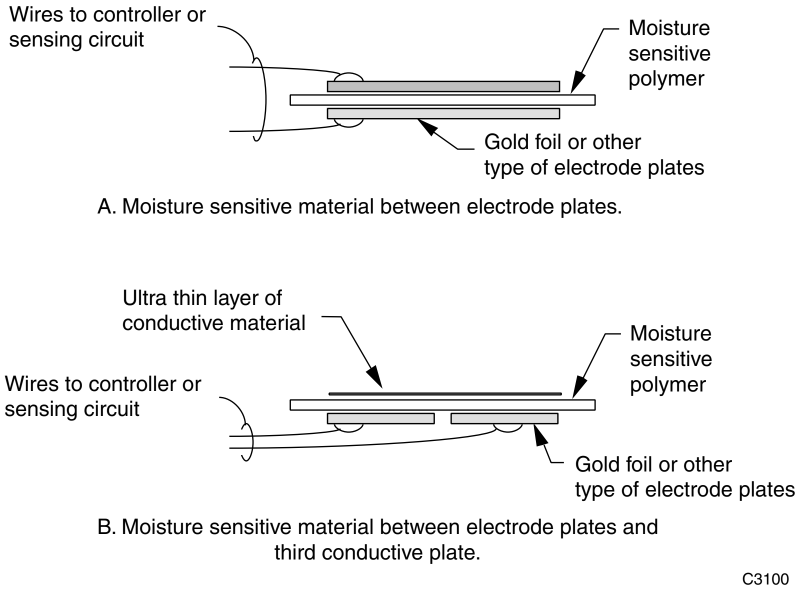


Fig. 10 Capacitance-type relative humidity sensor.

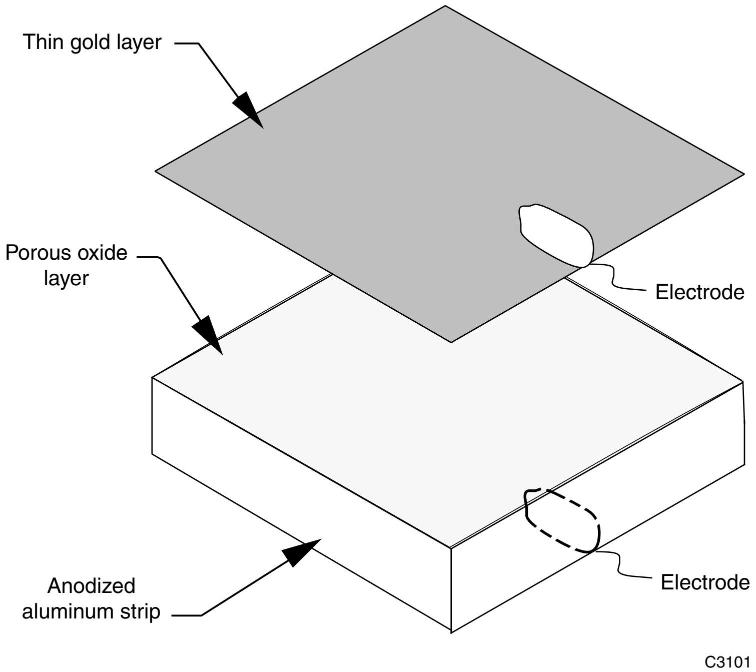


Fig. 11 Impedance-type relative humidity sensor.

sent to the controller is less susceptible to external noise interference. The sensor thus becomes a transmitter.

Another pressure sensing method measures capacitance (Fig. 14). A fixed plate forms one part of the capacitor assembly, and a flexible plate is the other part of the capacitor assembly. As the diaphragm flexes with pressure variations, the flexible plate of the capacitor assembly moves closer to the fixed plate (shown by a dotted line in Fig. 14) and changes the capacitance.

A variation of pressure sensors is one that measures differential pressure using dual pressure chambers (Fig. 15). The force from each chamber acts in an opposite direction with respect to the strain gage. This type of sensor can measure small differential pressure changes even with high static pressure.

Controllers, Output Devices and Indicating Devices

Controller

The electronic controller receives a sensor signal, amplifies and/or conditions it, compares it with the setpoint, and derives a correction if necessary. The output signal typically positions an actuator. Electronic controller circuits allow a wide variety of control functions and sequences, from very simple arrangements to multiple-input circuits with several sequential outputs. Controller circuits use solid-state components, such as transistors, diodes and integrated circuits, and include the power supply and all the adjustments required for proper control.

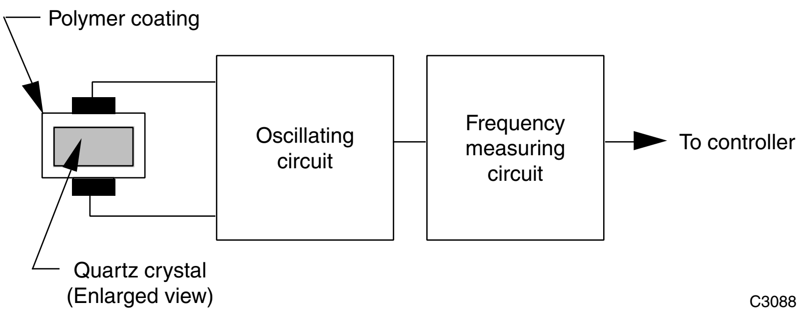


Fig. 12 Quartz crystal relative humidity sensor.

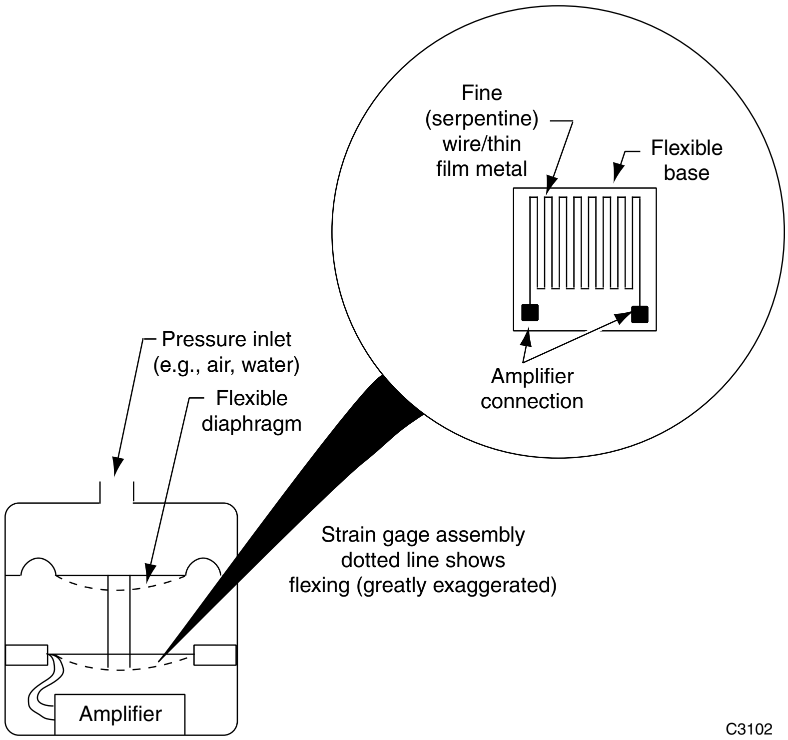


Fig. 13 Resistance-type pressure sensor.

Input Types

Electronic controllers are categorized by the type or types of inputs they accept, such as temperature, humidity, enthalpy or universal.

Temperature Controllers

Temperature controllers typically require a specific type or category of input sensors. Some have input circuits to accept RTD sensors such as BALCO or platinum elements, while others contain input circuits for thermistor sensors. These controllers have setpoint and throttling range scales labeled in degrees Fahrenheit or Celsius.

Relative Humidity Controllers

The input circuits for relative humidity controllers typically receive the sensed relative humidity signal already converted to a 0–10 V d.c. voltage, or a

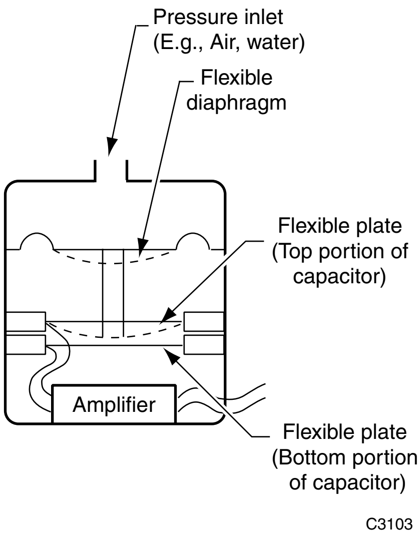


Fig. 14 Capacitance-type pressure transmitters.

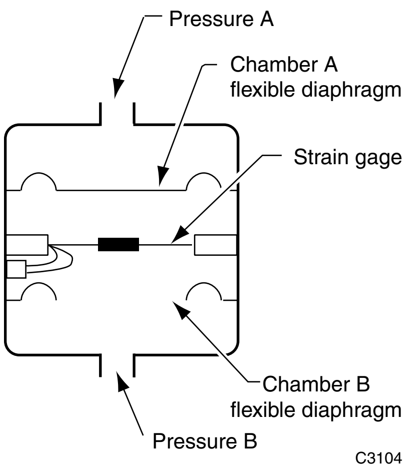


Fig. 15 Differential pressure sensor.

Elec–Elec

4–20 mA current signal. Setpoint and scales for these controllers are in percent relative humidity.

Enthalpy Controllers

Enthalpy controllers are specialized devices that use specific sensors for inputs. In some cases, the sensor may combine temperature and humidity measurements and convert them to a single voltage to represent enthalpy of the sensed air. In other cases, individual dry-bulb temperature sensors and separate wet-bulb or relative humidity sensors provide inputs, and the controller calculates enthalpy. In typical applications, the enthalpy controller provides an output signal based on a comparison of two enthalpy measurements, indoor and outdoor, rather than on the actual enthalpy value. In other cases, the return air enthalpy is assumed constant so that only OA enthalpy is measured. It is compared against the assumed nominal return air value.

Universal Controllers

The input circuits of universal controllers can accept one or more of the standard transmitter or transducer signals. The most common input ranges are 0–10 V d.c. and 4–20 mA. Other input variations in this category include a 2–10 V d.c. and a 0–20 mA signal. Because these inputs can represent a variety of sensed variables, such as a current of 0–15 A or pressure of 0–3000 psi, the settings and scales are often expressed in percent of full scale only.

Control Modes

The control modes of some electronic controllers can be selected to suit the application requirements. Control modes include two-position, proportional and proportional-integral. Other control features include remote setpoint, the addition of a compensation sensor for reset capability, and override or limit control.

Output Control

Electronic controllers provide outputs to a relay or actuator for the final control element. The output is not dependent on the input types or control method. The simplest form of output is two-position, in which the final control element can be in one of two states. For example, an exhaust fan in a mechanical room can be turned either on or off. The most common output form, however, provides a modulating output signal which can adjust the final control device (actuator) between 0 and 100 %, such as in the control of a chilled water valve.

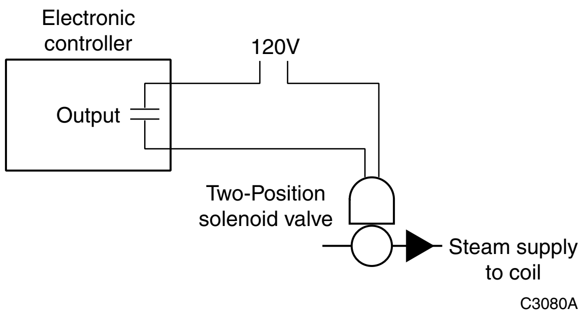


Fig. 16 Two-position control.

Output Devices

Actuators, relays, and transducers (Fig. 2) are output devices which use the controller output signal (voltage, current, or relay contact) to perform a physical function on the final control element such as starting a fan or modulating a valve. Actuators can be categorized as devices that provide two-position action or as those that provide modulating action.

Two-Position

Two-position devices such as relays, motor starters, and solenoid valves have only two discrete states. These devices interface between the controller and the final control element. For example, when a solenoid valve is energized, it allows steam to enter a coil that heats a room (Fig. 16). The solenoid valve provides the final action on the controlled media, steam. Damper actuators can also be designed to be two-position devices.

Modulating

Modulating actuators use a varying control signal to adjust the final control element. For example, a modulating valve controls the amount of chilled water entering a coil so that cool supply air is just sufficient to match the load at a desired setpoint (Fig. 17). The most common modulating actuators accept a varying voltage input of 0–10 V, or 2–10 V d.c., or a current input of 4–20 mA. Another form of actuator requires a pulsating (intermittent) or duty cycling signal to

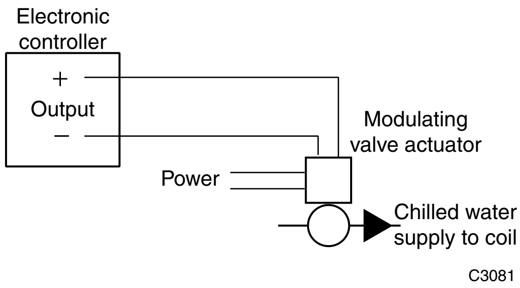


Fig. 17 Modulating control.

Elec–Elec

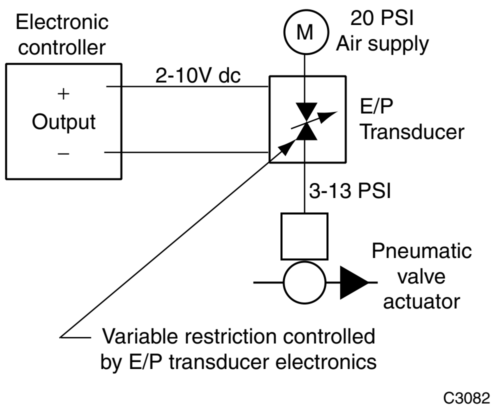


Fig. 18 Electric-to-pneumatic transducer.

perform modulating functions. One form of pulsating signal is a Pulse Width Modulation (PWM) signal.

Tranducer

In some applications, a transducer converts a controller output to a signal that is usable by the actuator. For example, Fig. 18 shows an Electronic-to-Pneumatic (E/P) transducer that converts a modulating 2–10 V d.c. signal from the electronic controller to a pneumatic proportional modulating 3–13 psi signal for a pneumatic actuator.

Indicating Devices

An electronic control system can be enhanced with visual displays that show system status and operation. Many electronic controllers have built-in indicators that show power, input signal, deviation signal and output signal. Fig. 19 shows some types of visual displays. An indicator light can show on/off status or, if driven by controller circuits, the brightness of a light can show the relative strength of a signal. If a system requires an analog or digital indicating device and the electronic controller does not include this type of display, separate indicating devices can be provided.

Interface with Other Systems

It is often necessary to interface an electronic control device to a microprocessor-based building management

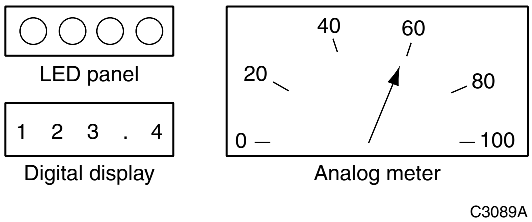


Fig. 19 Indicating devices.

system or other related system. An example is an interface that allows a building management system to adjust the setpoint or amount of reset (compensation) for a specific controller. Compatibility of the two systems must be verified before they are interconnected.

ELECTRONIC CONTROLLER FUNDAMENTALS

General

The electronic controller is the basis for an electronic control system. Fig. 20 shows the basic circuits of an electronic controller including power supply, input, control and output. For greater stability and control, internal feedback correction circuits also can be included, but these are not discussed here. The circuits described provide an overview of the types and methods of electronic controllers.

Power Supply Circuit

The power supply circuit of an electronic controller provides the required voltages to the input, control, and output circuits. Most voltages are regulated DC voltages. The controller design dictates the voltages and current levels required.

All power supply circuits are designed to optimize both line and load regulation requirements within the needs and constraints of the system. Load regulation refers to the ability of the power supply to maintain the voltage output at a constant value even as the current demand (load) changes. Similarly, line regulation refers to the ability of the power supply to maintain the output load voltage at a constant value when the input (AC) power varies. The line regulation abilities or limitations of a controller are usually part of the controller specifications such as 120 V AC +10%, –15%. The degree of load regulation involves the end-to-end accuracy and repeatability, and is usually not explicitly stated as a specification for controllers.

TYPICAL SYSTEM APPLICATIONS

Fig. 21 shows a typical air-handling system controlled by two electronic controllers, C1 and C2; sequencer S; multi-compensator M; temperature sensors T1 through T4; modulating hot- and chilled-water valves V1 and V2; and outdoor, return, and exhaust air damper actuators. The control sequence is as follows:

- Controller C1 provides outdoor compensated, summer/winter control of space temperature for a heating/cooling system which requires PI control with a low limit. Sensor T4 provides the compensation signal through multi-compensator M, which allows one

Elec–Elec

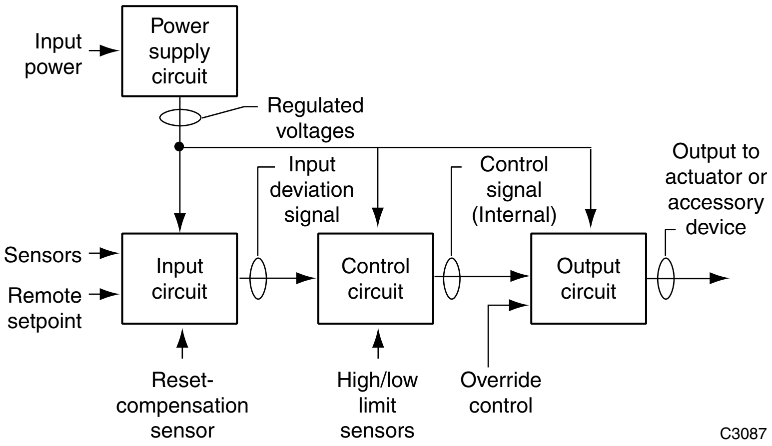


Fig. 20 Electronic controller circuits.

outdoor temperature sensor to provide a common input to several controllers. Controller C1 modulates the hot- and chilled-water valves V1 and V2 in sequence to maintain space temperature measured by sensor T1 at a pre-selected setpoint. Sequencer S allows sequencing the two valve actuators from a single controller. Low-limit sensor T2 assumes control when the discharge air temperature drops to the control range of the low-limit setpoint. A minimum discharge air temperature is maintained regardless of space temperature.

When the outdoor temperature is below the selected reset changeover point set on C1, the controller is in the winter compensation mode. As the outdoor air temperature falls, the space temperature setpoint is raised. When the outdoor temperature is above the reset changeover point, the controller is in the summer compensation mode. As the outdoor temperature rises, the space temperature setpoint is raised.

- Controller C2 provides PI mixed air temperature control with economizer operation. When the OA temperature measured by sensor T4 is below the setting of the economizer startpoint setting, the controller provides proportional control of the dampers to maintain mixed air temperature measured by

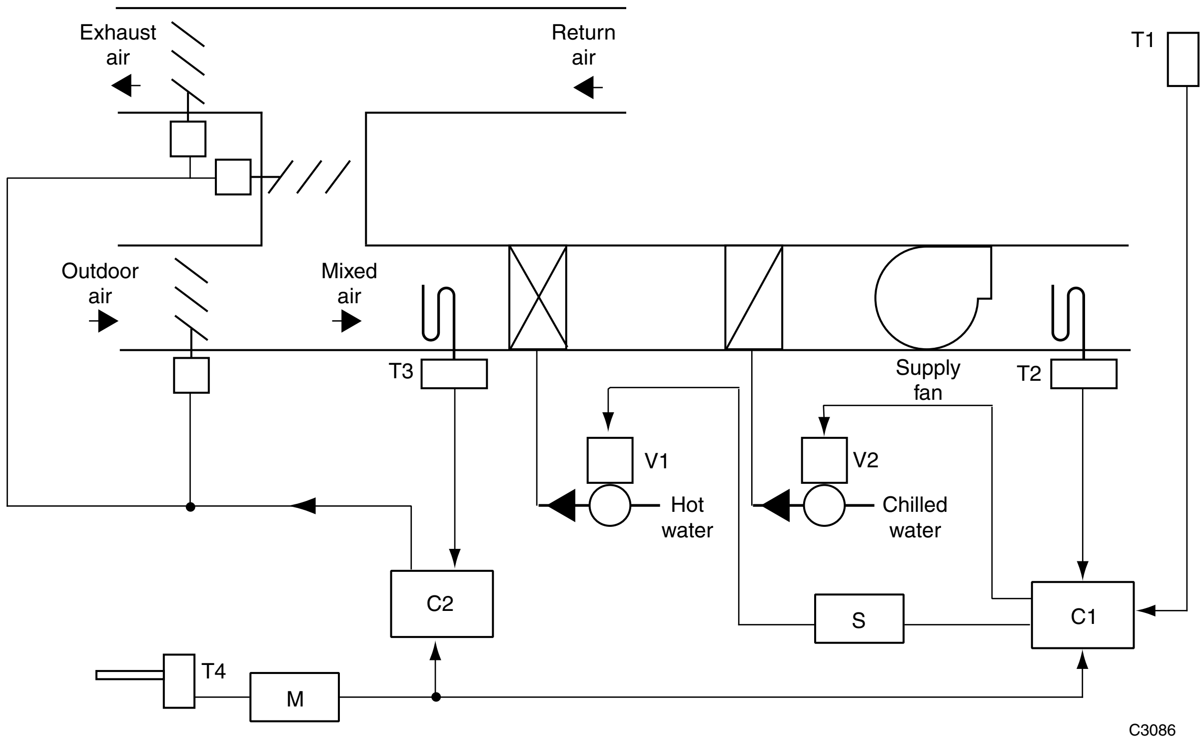


Fig. 21 Typical application with electronic controllers.

Elec–Elec

sensor T3 at the selected setpoint. When the OA temperature is above the economizer startpoint setting, the controller closes the OA dampers to a preset minimum.

ADDITIONAL DEFINITIONS

Authority (reset authority or compensation authority). A setting that indicates the relative effect a compensation sensor input has on the main setpoint (expressed in percent).

Compensation change-over. The point at which the compensation effect is reversed in action and changes from summer to winter or vice versa. The percent of compensation effect (authority) may also be changed at the same time.

Compensation control. See *Reset Control*.

Compensation sensor. See *Reset Sensor*.

Control Point. The actual value of a controlled variable (setpoint plus or minus offset).

Deviation. The difference between the setpoint and the value of the controlled variable at any moment. Also called “offset.”

Direct acting. A direct-acting controller increases its output signal on an increase in input signal.

Electric control. A control circuit that operates on line or low voltage, and uses a mechanical means, such as a temperature-sensitive bimetal or bellows, to perform control functions, such as actuating a switch or positioning a potentiometer. The controller signal usually operates or positions an electric actuator, although relays and switches are often controlled.

Electronic control. A control circuit that operates on low voltage and uses solid-state components to amplify input signals and perform control functions, such as operating a relay or providing an output signal to position an actuator. Electronic devices are primarily used as sensors. The controller usually furnishes fixed control routines based on the logic of the solid-state components.

Electronic controller. A solid-state device usually consisting of a power supply, a sensor amplification circuit, a process/comparing circuit, an output driver section, and various components that sense changes in the controlled variable and derive a control output which provides a specific control function. In general, adjustments such as setpoint and throttling range necessary for the process can be done at the controller via potentiometers and/or switches.

Final control element. A device such as a valve or damper that changes the value of the manipulated variable. The final control element is positioned by an actuator.

Integral action (I). An action in which there is a continuous linear relationship between the amount of increase (or decrease) on the output to the final control element and the deviation of the controlled variable to reduce or eliminate the deviation or offset.

Limit sensor. A device which senses a variable that may be other than the controlled variable and overrides the main sensor at a preset limit.

Main sensor. A device or component that measures the variable to be controlled.

Negative (reverse) reset. A compensating action in which a decrease in the compensation variable has the same effect as an increase in the controlled variable. For example, in a heating application, as the outdoor air temperature decreases, the control point of the controlled variable increases. Also called “winter reset or compensation.”

Offset. A sustained deviation between the control point and the setpoint of a proportional control system under stable operating conditions. Also called “deviation.”

Positive (direct) reset. A compensating action in which an increase in the compensation variable has the same effect as an increase in the controlled variable. For example, in a cooling application, as the OA temperature increases, the control point of the controlled variable increases. Also called “summer reset or compensation.”

Proportional band (throttling range). In a proportional controller, the control point range through which the controlled variable must pass to drive the final control element through its full operating range. Proportional band is expressed in percent of the main sensor span. A commonly used equivalent is “throttling range,” which is expressed in values of the controlled variable.

Proportional control (P). A control algorithm or method in which the final control element moves to a position proportional to the deviation of the value of the controlled variable from the setpoint.

Proportional-integral (PI) control. A control algorithm that combines the proportional (proportional response) and integral or deviation control algorithms. Integral action tends to correct the offset resulting from proportional control. Also called “proportional plus reset” or “two-mode” control.

Remote setpoint. A means for adjusting the controller setpoint from a remote location, in lieu of adjusting it at the controller itself. The means of adjustment may be manual with a panel or space mounted potentiometer, or automatic when a separate device provides a signal (voltage or resistive) to the controller.

Reset control. A process of automatically adjusting the control point of a given controller to compensate for changes in a second measured variable such as outdoor air temperature. For example, the hot deck control point is reset upward as the outdoor air temperature decreases. Also known as “compensation control.”

Reset sensor. The system element which senses a variable other than the controlled variable and resets the main sensor control point. The amount of this effect is established by the authority setting.

Reverse acting. A reverse-acting controller decreases its output signal on an increase in input signal.

Setpoint. The value on the controller scale at which the controller is set, such as the desired room temperature set on a thermostat. The setpoint is always referenced to the main sensor (not the reset sensor).

Throttling range. In a proportional controller, the control point range through which the controlled variable must pass to move the final control element through its full operating range. Throttling range is expressed in values of the controlled variable such as temperature in degrees Fahrenheit, relative humidity in percent, or pressure in pounds per square inch. A commonly used equivalent is “proportional band,” which is expressed in percent of sensor span for electronic controls.

Transducer. A device that converts one energy form to another. It amplifies (or reduces) a signal so that the output of a sensor or transducer is usable as an input to a controller or actuator. A transducer can convert a pneumatic signal to an electric signal (P/E transducer) or vice versa (E/P transducer), or it can convert a change in capacitance to an electrical signal.

Transmitter. A device that converts a sensor signal to an input signal usable by a controller or display device.

CONCLUSION

Basic automatic electronic control systems are extremely important to provide desirable operational features of energy-using equipment and systems. Proper control is critical to achieving functional performance, as well as energy-efficient performance in equipment, buildings and processes. Electronic control systems hold a large share of the control technologies used in most of our modern energy control applications.

REFERENCES

1. Engineering Manual of Automatic Controls, Honeywell International, Golden Valley, MN, 2005; available online at www.buildingcontrolworkbench.com.

BIBLIOGRAPHY

1. Control Systems. In *Chapter 12 in Energy Management Handbook*; 5th Ed.; Turner, W.C., Ed.; Fairmont Press Atlanta, GA, 2005.
2. Control Systems and Computers. In *Chapter 9 in Guide to Energy Management*; 5th Ed.; Capehart, B.L., Turner, W.C., and Kennedy, W.J., Eds.; Fairmont Press Atlanta, GA, 2006.
3. HVAC Control Systems, ASHRAE Handbook of Fundamentas In *Chapter 15*, *American Society of Heating, Refrigerating, and Air Conditioning Engineers*, Atlanta, GA.
4. Wolovich, W.A. In *Automatic Control Systems: Basic Analysis and Design*; Oxford University Press: U.S.A., 1994.
5. Bradshaw, V. In *Building Control Systems*; Wiley: New York, 1993.
6. Kuo, B.C. *Automatic Control Systems*, 7th Ed.; Wiley: New York, 1995.
7. Underwood, C. *HVAC Control Systems*; Taylor and Francis Books, Ltd.: Washington, DC, 1998.
8. Bolton, W. *Control Engineering*, 2nd Ed.; Pearson Higher Education: 1998.
9. Bishop, R.H.; Dorf, R.C. *Modern Control Systems*, 9th Ed.; Pearson Higher Education: 2000.

Emergy Accounting

Mark T. Brown
Department of Environmental Engineering Sciences, University of Florida, Gainesville, Florida, U.S.A.

Sergio Ulgiati
Department of Sciences for the Environment, Parthenope, University of Naples, Napoli, Italy

Abstract
In this chapter, we briefly review H.T. Odum's concepts and principles of emergy and related quantities.[1–5] The concept of energy quality is introduced and defined by transformity and specific emergy. Tables are given of data on global emergy flows, from which the emergy and transformities of most products and processes of the biosphere are calculated. Tables of transformity and specific emergy for many secondary products are provided. Finally, the concept of net emergy yield is introduced and defined using an Emergy Yield Ratio (EYR).

DEFINITIONS

Energy is sometimes referred to as the ability to do work. Energy is a property of all things that can be turned into heat and is measured in heat units (Btus, calories, or joules).

Emergy is the availability of energy [exergy (See the entry "Exergy" in this same encyclopedia.) For high-quality flows, such as fuels and electricity, the energy content and the available energy do not differ significantly. For this reason, the energy of a flow instead of its exergy is sometimes used for the sake of simplicity.] of one kind that is used up in transformations directly and indirectly to make a product or service. The unit of emergy is the emjoule, a unit referring to the available energy of one kind consumed in transformations. For example, sunlight, fuel, electricity, and human service can be put on a common basis by expressing them all in the emjoules of solar energy that are required to produce each. In this case, the value is a unit of solar emergy expressed in solar emjoules (abbreviated sej). Although other units have been used, such as coal emjoules or electrical emjoules, in most cases all emergy data are given in solar emjoules.

The emjoule, short for "emergy joule," is the unit of measure of emergy. It is expressed in the units of the energy previously used to generate the product; the solar emergy of wood, for example, is expressed in the joules of solar energy that were required to produce the wood.

The emdollar (abbreviated em$) is a measure of the money that circulates in an economy as the result of some process. In practice, to obtain the emdollar value of an emergy flow or storage, the emergy is multiplied by the ratio of total emergy to the Gross National Product for the national economy.

Unit emergy values are calculated based on the emergy required to produce them. There are three types of unit emergy values, as follows:

Transformity is defined as the emergy per unit of available energy (exergy). For example, if 4000 solar emjoules are required to generate a joule of wood, the solar transformity of that wood is 4000 solar emjoules per joule (abbreviated sej/J). Solar energy is the largest, but most dispersed, energy input to the Earth. The solar transformity of the sunlight absorbed by the Earth is 1.0 by definition.

Specific emergy is the unit emergy value of matter defined as the emergy per mass, usually expressed as solar emergy per gram (abbreviated sej/g). Solids may be evaluated best with data on emergy per unit mass for its concentration. Because energy is required to concentrate materials, the unit emergy value of any substance increases with concentration. Elements and compounds not abundant in nature therefore have higher emergy/mass ratios when found in concentrated form, because more work was required to concentrate them, both spatially and chemically.

Emergy per unit money is a unit emergy value used to convert money payments to emergy units. The amount of resources that money buys depends on the amount of emergy supporting the economy and the amount of money circulating. An average emergy/money ratio in solar emjoules per dollar can be calculated by dividing the total emergy use of a state or nation by its gross economic product. It varies by country and has been shown to decrease each year. This emergy/money ratio is useful for evaluating service inputs given in money units when an average wage rate is appropriate.

Keywords: Energy quality; Emergy; Emergy yield ratio; Specific emergy; Transformity.

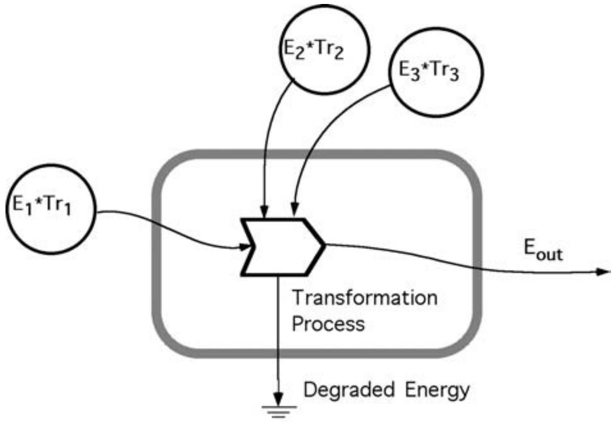
Empower is a flow of emergy (i.e., emergy per unit of time). Emergy flows are usually expressed in units of solar empower (solar emjoules per unit of time).

ENERGY, QUALITY, AND EMERGY

Probably the least understood and most criticized parts of H.T. Odum's body of work[1–5] are his concepts and theories of energy quality, which are embodied in the 35 year development of the emergy concept. The development of emergy and its theoretical base cannot be separated from the development of the concept of energy quality. Beginning in the early 1970s, Odum suggested that different forms of energy have different capacities to do work. He reasoned that whereas energy is measured in units of heat (Btus, joules, calories), not all calories are the same when it comes to work processes, especially complex work processes. All energies can be converted to heat at 100% efficiency; thus, it is relatively easy and accurate to express energies in their heat equivalents.

Although heat-equivalent energy is a good measure of the ability to raise the temperature of water, it is not a good measure of more complex work processes. Processes outside the window defined by heat engine technology do not use energies that lend themselves to thermodynamic heat transfers. As a result, converting all energies of the biosphere to their heat equivalents reduces all work processes of the biosphere to heat engines. Human beings, then, become heat engines, and the value of their services and information is nothing more than a few thousand calories per day. Different forms of energy have different abilities to do work. A calorie of sunlight is not the same as a calorie of fossil fuel or a calorie of food unless it is being used to power a steam engine. A system organized to use concentrated energies like fossil fuels (or food) cannot process a more dilute energy form like sunlight, calorie for calorie. By the same token, a system organized to process dilute energy like sunlight (a plant, for instance) cannot process more concentrated energy like fossil fuels.

In this way, the use and transformation of energy sources is system dependent; the appropriateness of an energy source in a particular system is dictated by its form and is related to its concentration. The processes of the biosphere are infinitely varied and are more than just thermodynamic heat engines. As a result, the use of heat measures of energy that can recognize only one aspect of energy—its ability to raise the temperature of things—cannot quantify adequately the work potential of energies used in more complex processes of the biosphere. In the larger biosphere system as a whole, energies should be converted to units that span this greater realm, accounting for multiple levels of system processes, ranging from the smallest scale to the largest scales of the biosphere, and accounting for processes other than heat engine technology.



$$Em_{out} = \sum E_n{*}Tr_n$$

$$Tr = Em_{out} / E_{out}$$

Where;

$E_{1...n}$ = Available energy inputs
E_{out} = Available energy of output
Em = Emergy
Tr = Transformity

Fig. 1 In all processes, some energy is degraded and some is transformed into higher quality energy. The energy out is equal to the sum of the input energies minus the degraded energy. The emergy out is equal to the sum of the input emergies. The equations at the bottom of the figure show the general calculation of emergy of a product.

TRANSFORMITY AND SPECIFIC EMERGY

Transformity and specific emergy are unit emergy values calculated as the total amount of emergy required to make a product or service divided by the available energy of the product (resulting in a transformity) or divided by the mass of the product (resulting in a specific emergy). Figs. 1 and 2 illustrate the method of calculating a transformity, first in equation form (Fig. 1) and then with example numbers (Fig. 2). The transformity of the product is the emergy of the product divided by the energy of the product (in sej/J). If the output flow is in mass, the specific emergy of the product is the emergy of the output divided by the mass (in sej/g).

TRANSFORMITY AND QUALITY

Quality is a system property, which means that an "absolute" scale of quality cannot be made; neither can the usefulness of a measure of quality be assessed without first defining the structure and boundaries of the system. Self-organizing systems (be they the biosphere or a national economy) are organized with hierarchical levels (Fig. 3), and each level is composed of many parallel processes. This leads to two possible definitions of quality: parallel quality and cross quality. The first, parallel

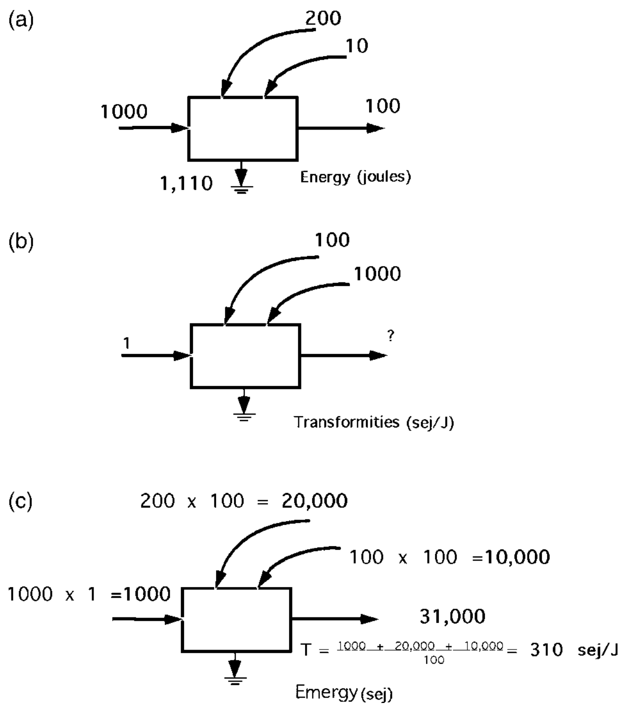


Fig. 2 Method of calculating transformity. (a) energy flows; (b) transformity of the output is calculated by dividing the emergy of the output in c by the energy of the output in a.

quality, is related to the efficiency of a process that produces a given flow of energy or matter within the same hierarchical level. (See Fig. 3 for an example of comparison among units in the same hierarchical level.) For any given output—say, electricity—there is almost an infinite number of ways to produce it, including all the generators, chemical processes, solar voltaic cells, and hydroelectric dams presently in service. A recent compilation of transformities for electricity from various production systems has yielded transformities from 6.23×10^4 sej/J, for a 2.5-MW wind generator, to 2.0×10^5 sej/J,

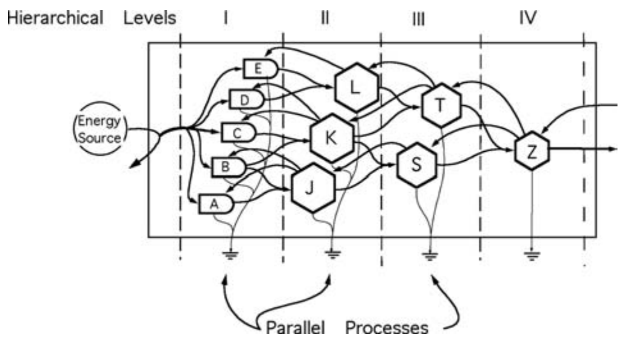


Fig. 3 Complex systems are organized hierarchically where it takes many small components to support the next level in the hierarchy which in turn supports even fewer components at the next level, and so on. Comparison between components of the same level is a comparison of parallel quality while comparison of components from different levels is a comparison of cross quality.

for a 1280-MW oil-fired power plant in Italy.[6] A mean value of 1.66×10^5 sej/J from these and other plants was suggested by Odum for electricity when the source is assumed to come from "average" conditions.[4] The same rationale can be used for any energy or material flow as long as the flow is believed to represent the average. Each individual process has its own efficiency, and as a result, the output has a distinct transformity. Quality as measured by transformity in this case relates to the emergy required to make like products under differing conditions and processes. For the most part, transformities of like products are within the same order of magnitude.

The second definition of quality, cross quality, is related to the hierarchical organization of systems. In this case, transformity is used to compare components or outputs from different levels of the hierarchy, accounting for the convergence of emergy at higher and higher levels. (See Fig. 3 for an example of comparison of transformity between different hierarchical levels.) At higher levels, a larger convergence of inputs is required to support the component: many villages are necessary to support a city, many kilograms of grass to support a cow, etc. Also, higher feedback and control ability characterize components at higher hierarchical levels. Therefore, higher transformity, as equated with a higher level in the hierarchy, often means greater flexibility and is accompanied by greater spatial and temporal effects. In this definition of quality, the higher the transformity, the higher the quality of the process or product. Transformities of products from different hierarchical levels usually differ by at least one order of magnitude.

EMERGY OF THE GEOBIOSPHERE: THE BASIS FOR COMPUTING UNIT EMERGY VALUES

Annual Budget of Emergy Supporting the Geobiosphere

An emergy evaluation table of the main inputs to the geobiosphere of the Earth (omitting, for the moment, the emergy use from nonrenewable resources) is given in Table 1. The annual budget of emergy flow (empower) supporting the geobiosphere (the atmosphere, ocean, and Earth's crust) includes solar energy, tidal energy, and heat energy from the deep Earth. These contributions to the geobiosphere total about 15.83×10^{24} sej/year.

Average Emergy Unit Values for Main Global Processes

Table 2 calculates unit emergy values for some main flows of the Earth. The total emergy input to the geobiosphere in solar emergy (15.83×10^{24} sej/year from Table 1) is divided by each global product's ordinary measure (the number of joules or grams). The unit values that result are

Table 1 Annual emergy contributions to global processes

Input	Inflow (J/year)	Emergy (sej/J)	Empower $\times 10^{24}$ (sej/year)
Solar insolation[a]	3.93×10^{24}	1.0	3.93
Deep earth heat[b]	6.72×10^{20}	1.20×10^{4}	8.06
Tidal energy[c]	0.52×10^{20}	7.39×10^{4}	3.84
Total	—	—	15.83

Not including non-renewable resources.
Abbreviations: sej, solar emjoules.
[a]Sunlight: solar constant 2 gcal/cm^2/min = 2 Langley per minute; 70% absorbed; earth cross section facing sun 1.27×10^{14} m^2.
[b]Heat release by crustal radioactivity 1.98×10^{20} J/year plus 4.74×10^{20} J/year heat flowing up from the mantle Ref. 16. Solar transformity 1.2×10^{4} sej/J based on an emergy equation for crustal heat as the sum of emergy from earth heat, solar input to earth cycles, and tide Ref. 8.
[c]Tidal contribution to oceanic geopotential flux is 0.52×10^{20} J/year Ref. 17. Solar transformity of 7.4×10^{4} sej/J is based on an emergy equation for oceanic geopotential as the sum of emergy from earth heat, solar input to the ocean, and tide following Refs. 8 and 18.
Source: From University of Florida (see Ref. 7).

useful for other emergy evaluations for which global averages can be used.

Temporary Emergy Inputs to the Geobiosphere

In the past two centuries, the production and consumption processes of human civilization that are using the large emergy in the geologic stores of fuels and minerals have reached a scale with global impact. Because these storages are being used much faster than they are being generated in geologic cycles, they are often called nonrenewable resources. They are actually very slowly renewed resources. Table 3 summarizes these additional components of the global emergy budget.

At present, the emergy from nonrenewable energy use that is contributed to the geobiosphere by human civilization is greater than the inputs from renewable sources. The result of this "temporary surge" of emergy is the accumulation of carbon dioxide in the atmosphere, adding to the greenhouse effects that may be altering ocean temperatures and, ultimately, the pattern and intensity of weather. The total renewable and nonrenewable emergy contributions to the global systems, including those released by humans, are 50.1×10^{24} sej/year. (Fig. 4).

UNIT EMERGY VALUES FOR FUELS AND SOME COMMON PRODUCTS

Unit emergy values result from emergy evaluations. Following are several tables of unit emergy values for some common materials and energy sources. In Table 4,

Table 2 Emergy of products of the global energy system

Product	Emergy[a] $\times 10^{24}$ (sej/year)	Production	Emergy
Global latent heat[b]	15.83	1.26×10^{24} J/year	12.6 sej/J
Global wind circulation[c]	15.83	6.45×10^{21} J/year	2.5×10^{3} sej/J
Global precipitation on land[d]	15.83	1.09×10^{20} g/year	1.5×10^{5} sej/g
Global precipitation on land[e]	15.83	5.19×10^{20} J/year	3.1×10^{4} sej/J
Average river flow[f]	15.83	3.96×10^{19} g/year	4.0×10^{5} sej/g
Average river geopotential[g]	15.83	3.4×10^{20} J/year	4.7×10^{4} sej/J
Average river chem. energy[h]	15.83	1.96×10^{20} J/year	8.1×10^{4} sej/J
Average waves at the shore[i]	15.83	3.1×10^{20} J/year	5.1×10^{4} sej/J
Average ocean current[j]	15.83	8.6×10^{17} J/year	1.8×10^{7} sej/J

[a]Main empower of inputs to the geobiospheric system from Table 1 not including non-renewable consumption (fossil fuel and mineral use).
[b]Global latent heat = latent heat of evapotranspiration 1020 mm/year, (1020 mm/year)(1000 g/m^2/mm)(0.58 Cal/g)(4186 J/Cal)(5.1×10^{14} m^2) = 1.26×10^{24} J/year.
[c]Global wind circulation, 0.4 watts/m^2 Ref. 19 (0.4 J/m^2/sec)(3.15×10^{7} sec/year)(5.12×10^{14} m^2/earth) = 6.45×10^{21} J/year.
[d]Global precipitation on land = 1.09×10^{11} m^3/year Ref. 20 (1.09×10^{14} m^3)(1×10^{6} kg/m^3) = 1.09×10^{20} g/year.
[e]Chemical potential energy of rain water relative to sea water salinity (1.09×10^{20} g/year)(4.94 J Gibbs free energy/g) = 5.19×10^{20} J/year.
[f]Global runoff, 39.6×10^{3} km^3/year (Todd 1970) (39.6×10^{12} m^3/year)(1×10^{6} g/m^3) = 3.96×10^{19} g/year.
[g]Average river geopotential work; average elevation of land = 875 m (39.6×10^{12} m^3/year)(1000 kg/m^3)(9.8 m/sec^2)(875 m) = 3.4×10^{20} J/year.
[h]Chemical potential energy of river water relative to sea water salinity (3.96×10^{19} g/year)(4.94 J Gibbs free energy/g) = 1.96×10^{20} J/year.
[i]Average wave energy reaching shores, Ref. 21 (1.68×10^{8} Cal/m/year)(4.39×10^{8} m shore front)(4186 J/Cal) = 3.1×10^{20} J/year.
[j]Average ocean current: 5 cm/sec Oort et al. 1989; 2 year turnover time (0.5)(1.37×10^{21} kg water)(0.050 m/sec)(0.050 m/sec)/(2 year) = 8.56×10^{17} J/year.
Source: From University of Florida (see Ref. 7).

Table 3 Annual emergy contributions to global processes including use of resource reserves

Inputs	Inflow (J/year)	Emergy[a]	Empower $\times 10^{24}$ (sej/year)
Renewable inputs[b]	—	—	15.8
Nonrenwable energies released by society:			
Oil[c]	1.38×10^{20}	9.06×10^{4} sej/J	12.5
Natural gas (oil eq.)[d]	7.89×10^{19}	8.05×10^{4} sej/J	6.4
Coal (oil eq.)[e]	1.09×10^{20}	6.71×10^{4} sej/J	7.3
Nuclear power[f]	8.60×10^{18}	3.35×10^{5} sej/J	2.9
Wood[g]	5.86×10^{19}	1.84×10^{4} sej/J	1.1
Soils[h]	1.38×10^{19}	1.24×10^{5} sej/J	1.7
Phosphate[i]	4.77×10^{16}	1.29×10^{7} sej/J	0.6
Limestone[j]	7.33×10^{16}	2.72×10^{6} sej/J	0.2
Metal ores[k]	9.93×10^{14}	1.68×10^{9} sej/g	1.7
Total non-renewable empower			34.3
Total global empower			50.1

Abbreviations: sej, solar emjoules; t, metric ton; oil eq., oil equivalents.
[a]Values of solar emergy/unit from Ref. 4 and modified to reflect a global resource base of 15.83×10^{24} sej/year.
[b]Renewable Inputs: Total of solar, tidal, and deep heat empower inputs from Ref. 4.
[c]Total oil production $= 3.3 \times 10^{9}$ t oil equivalent Ref. 8. Energy flux $= (3.3 \times 10^{9}$ t oil eq.)$(4.186 \times 10^{10}$ J/t oil eq.) $= 1.38 \times 10^{20}$ J/year oil equivalent.
[d]Total natural gas production $= 2.093 \times 10^{9}$ m^3 Ref. 8. Energy flux $= (2.093 \times 10^{12}$ $m^3)(3.77 \times 10^{7}$ J $m^3) = 7.89 \times 10^{19}$ J/year.
[e]Total soft coal production $= 1.224 \times 10^{9}$ t/year Ref. 8. Total hard coal production $= 3.297 \times 10^{9}$ t/year Ref. 8. Energy flux $= (1.224 \times 10^{9}$ t/year)$(13.9 \times 10^{9}$ J/t) $+ (3.297 \times 10^{9}$ t/year)$(27.9 \times 10^{9}$ J/t) $= 1.09 \times 10^{20}$ J/year.
[f]Total nuclear power production $= 2.39 \times 10^{12}$ kwh/year Ref. 8. Energy flux $= (2.39 \times 10^{12}$ kwh/year)$(3.6 \times 10^{6}$ J/kwh) $= 8.6 \times 10^{18}$ J/year electrical equivalent.
[g]Annual net loss of forest area $= 11.27 \times 10^{6}$ ha/year Ref. 23. Biomass $= 40$ kg m^2; 30% moisture Ref. 24. Energy flux $= (11.27 \times 10^{6}$ ha/year)$(1 \times 10^{4}$ m^2/ha)(40 kg m^2)$(1.3 \times 10^{7}$ J/kg)(0.7) $= 5.86 \times 10^{19}$ J/year.
[h]Total soil erosion $= 6.1 \times 10^{10}$ t/year Refs. 25 and 26. Assume soil loss 10 t/ha/year and 6.1×10^{9} ha agricultural land $= 6.1 \times 10^{16}$/g/year (assume 1.0% organic matter), 5.4 Cal/g. Energy flux $= (6.1 \times 10^{16}$ g)(0.01)(5.4 Cal/g)(4186 J/Cal) $= 1.38 \times 10^{19}$ J/year.
[i]Total global phosphate production $= 137 \times 10^{6}$ t/year Ref. 27. Gibbs free energy of phosphate rock $= 3.48 \times 10^{2}$ J/g. Energy flux $= (137 \times 10^{12}$ g)$(3.48 \times 10^{2}$ J/g) $= 4.77 \times 10^{16}$ J/year.
[j]Total limestone production $= 120 \times 10^{6}$ t/year Ref. 27. Gibbs free energy phosphate rock $= 611$ J/g. Energy flux $= (120 \times 10^{12}$ g)$(6.11 \times 10^{2}$ J/g) $= 7.33 \times 10^{16}$ J/year.
[k]Total global production of metals 1994: Al, Cu, Pb, Fe, Zn Ref. 28: 992.9×10^{6} t/year $= 992.9 \times 10^{12}$ g/year.
Source: After Elsevier (see Ref. 30).

unit emergy values are given for primary nonrenewable energy sources. In some cases, the unit emergy value is based on only one evaluation—plantation pine, for example. In other cases, several evaluations have been done of the same primary energy but from different sources, and presumably different technology, so unit emergy value is an average. Obviously, each primary energy source has a range of values depending on source and technology. Because they use data from typical production facilities (and actual operating facilities), the unit emergy values represent average conditions and can be used for evaluations when actual unit values are not known. If it is known that the conditions under which an evaluation is being conducted are quite different from the averages suggested here, detailed evaluations of the sources should be conducted.

Table 5 lists the unit emergy values for some common products in order of their transformity. Only a few products are given in this table, but many more evaluations leading to unit emergy values have been conducted and are presented in a set of Emergy Folios published by the Center for Environmental Policy at the University of Florida.[7,8]

NET ENERGY AND EMERGY YIELD RATIO

The concept of net energy has played an important role in the development of energy quality and emergy. The concept of net production in ecosystems is widely used and understood as a measure of fitness of ecological systems. When applied to the human economy, the concept suggests that an energy source must be able to provide a net contribution to the economy of the larger system in which it is embedded (i.e., it must provide more energy than it costs in extraction and processing). Because this principle must be applicable to all living systems—from ATP-providing energy to the biochemical reactions in

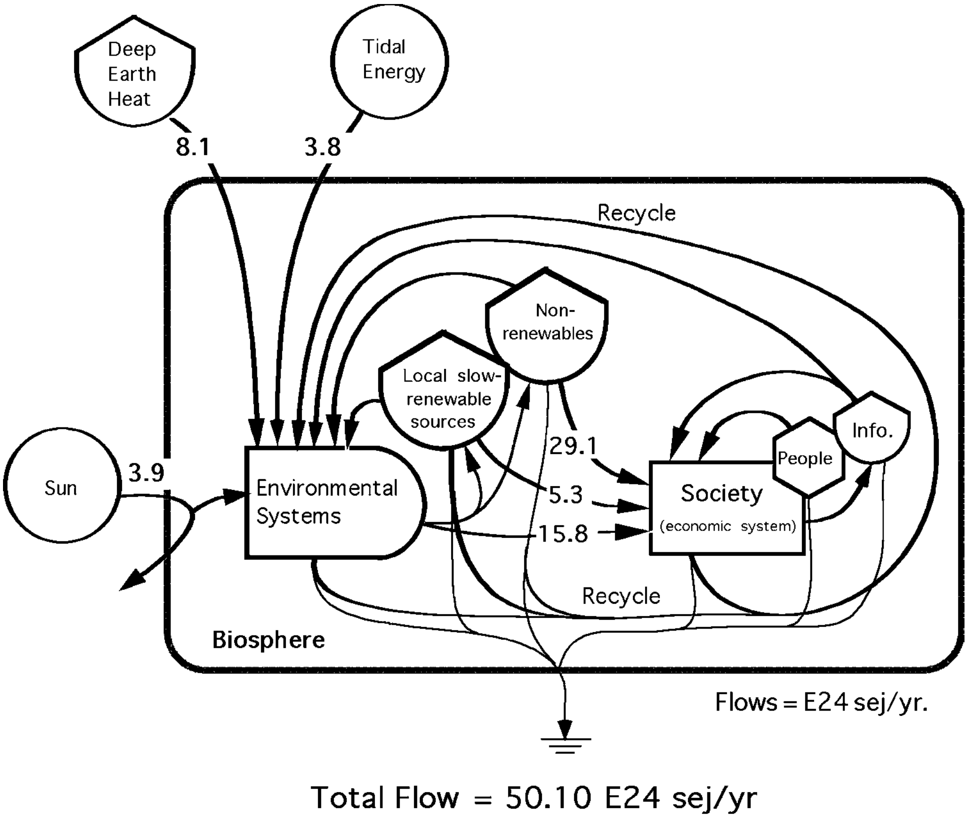


Fig. 4 Total emergy driving biosphere processes. For millions of years the driving emergy was dominated by the sources of sun, deep heat, and tidal momentum, now the dominant emergy flows are those associated with human society.

living systems, to photosynthesis, to the energy expended by animals as they graze or chase prey—it seems logical that it must also be applied to the processes of extracting fossil fuels from the Earth and to energy production of all sorts that drives economic sectors and human societies.

Table 4 Unit emergy values for primary non-renewable energy sources

	Transformity	
Item	**sej/J**	**sej/g**
Plantation pine (in situ)	1.1×10^4	9.4×10^7
Peat	3.2×10^4	6.7×10^8
Lignite	6.2×10^4	
Coal	6.7×10^4	
Rainforest wood (chipped, trans.)	6.9×10^4	4.1×10^8
Natural gas	8.1×10^4	
Crude oil	9.1×10^4	
Liquid motor fuel	1.1×10^5	
Electricity	3.4×10^5	

Source: From Wiley (see Ref. 4).

Odum suggested, "The true value of energy to society is the net energy, which is that after the costs of getting and concentrating that energy are subtracted."[9]

An Emergy Yield Ratio (EYR) is used to calculate the net contribution of energy sources to the economy. The EYR, as its name implies, is the ratio of the yield from a process (in emergy) to its costs (in emergy). The diagram in Fig. 5 illustrates the concept. The yield from this process is the sum of the input emergy from all sources: the environmental renewable source on the left (R), the nonrenewable storage (N), and the two purchased flows from the right (F). In the case of fossil fuels, there is little input from renewable sources, because the vast majority of the input comes from deep storages in the Earth. The EYR is the ratio of the yield (Y) to the costs (F) of retrieving it. The costs include energy, materials, and human service purchased from the economy, all expressed in emergy.

In practice, calculating the net energy of a process is far more complex than is shown in Fig. 5. Most processes, especially energy technologies, have many stages (or unit processes) and many inputs at each stage. As an example, Fig. 6 illustrates a series of processes beginning with an "energy crop"—fast-growing willow—and ending with wood chips that will be used in an electric power plant. At each stage in the process, an EYR can be calculated.

Table 5 Unit emergy values for some common products

Item	Transformity (sej/J)	Specific emergy (sej/g)
Corn stalks	6.6×10^4	
Rice, high energy[a]	7.4×10^4	1.4×10^9
Cotton	1.4×10^5	
Sugar (sugar cane)[b]	1.5×10^5	
Corn	1.6×10^5	2.4×10^9
Butter	2.2×10^6	
Ammonia fertilizer	3.1×10^6	
Mutton	5.7×10^6	
Silk	6.7×10^6	
Wool	7.4×10^6	
Phosphate fertilizer	1.7×10^7	
Shrimp (aquaculture)	2.2×10^7	
Steel[b]	8.7×10^7	7.8×10^9

[a]After Ref. 22.
[b]After Ref. 29.
Source: From Wiley (see Ref. 4).

Notice that the EYR decreases at each stage as more and more resources are used to process the wood into chips.

NET EMERGY AND TIME

Net energy is related to time, in that the longer a product has to develop, the higher its quality and the greater its net contribution. Doherty, for example, evaluated several processes that convert wood to higher-quality energy, such as ethanol.[10] Then he graphed the EYR vs the time it takes to grow the input wood. The graph in Fig. 7 resulted. The cycle time in this case is the time (in years) it takes to grow the wood from seeding to harvest. Some processes use fast-growing shrubby biomass (willow), whereas the longer cycle times are for climax rainforest species that take well over 100 years to mature. As the time increases, the quality of the wood (heat content) increases; thus, the EYR increases.

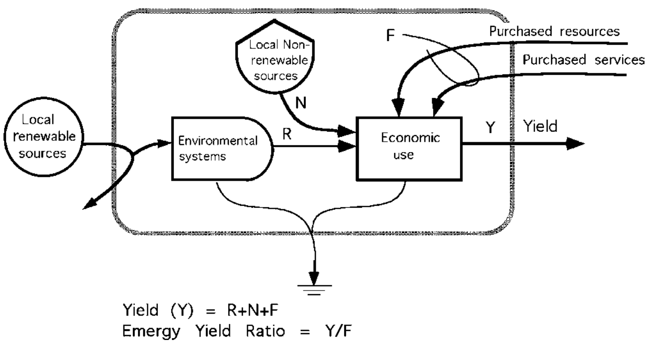


Fig. 5 Emergy Yield Ratio (EYR) is equal to the yield (expressed in emergy) divided by the purchased resources and services (also expressed in emergy).

Corn converted to ethanol is sometimes offered as an alternative fuel to fossil fuels. The impressive increases in yields per acre over time have been offered as evidence of the potential for corn to provide net energy and ultimately provide a renewable alternative to conventional fuels. The corn-to-ethanol process, however, uses a significant amount of fossil fuels directly and indirectly in growing, transporting, and conversion processes. In addition, the net yields are such that the result of using ethanol made from corn is that more energy will be consumed. Finally, there is no evidence that the net energy is increasing; therefore, the contribution to the economy is not increasing. Fig. 8 shows a graph of EYR and transformity for U.S. corn production from 1945 to 1993.[11–15] The EYR appears to be relatively constant at a value of around 1.7 over the past 55 years, which means that the contribution to the economy is not changing, even with increased yields. Transformities show declining values over this same period—probably the result of increased efficiency in the use of input resources and the fact that there have been increasing yields per hectare. As the EYR is constant, however, the net yield to society is not increasing.

NET EMERGY OF ENERGY SOURCES

Critical to continued prosperity, the net yields from fossil fuel energy sources that drive our economy are declining. As the richest and largest oil fields are tapped, and the remaining energy gets harder to find and even harder to drill for, the energy costs of obtaining oil and gas rise. As these limits are felt throughout modern economies, society looks to alternative sources: wind, waves, tides, solar, biomass, ethanol, and others. The graphs in Figs. 9 and 10 show the EYR for various energy sources used in modern economies. In Fig. 9, conventional nonrenewable sources are shown, and in Fig. 10, some of the so-called renewable energies are shown. It is imperative that the net contributions of proposed new energy sources be evaluated and all costs included. Many of the so-called renewable energy sources are actually guzzlers of fossil fuels. Take, for instance, proposed corn-to-ethanol programs. Our repeated evaluations over the past decade continue to show net yields of less than 2–1. Therefore, if ethanol is used to replace fossil fuels having yields of 8–1, the ethanol is actually using energy at four times the rate and increasing greenhouse-gas emissions over the burning of fossil fuels.

CONCLUSION

Emergy accounting—the process of evaluating the contributions of energy, material, and information inputs to processes—is intended to account for aspects that are

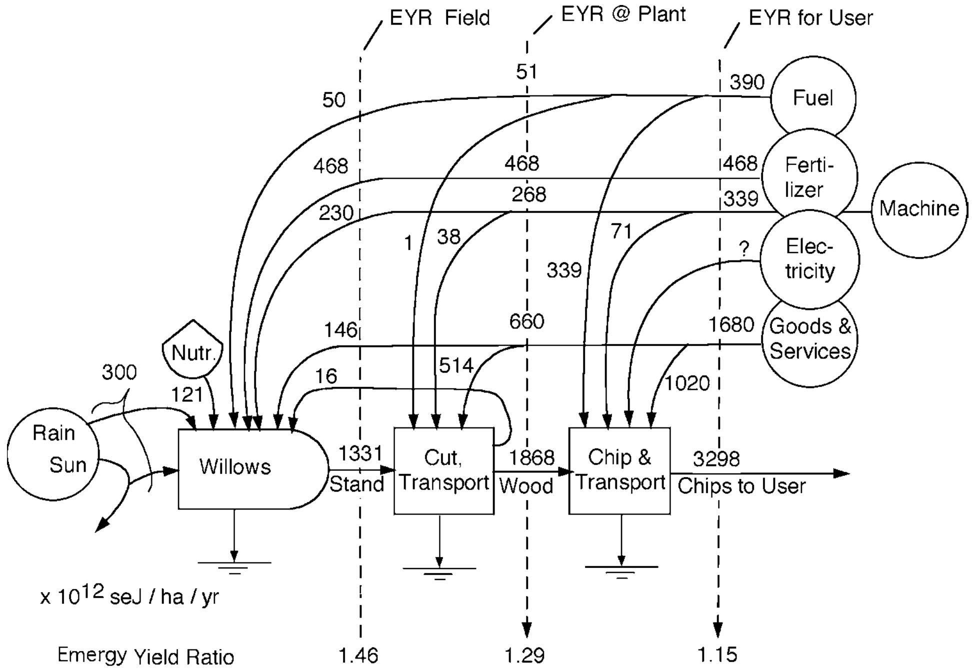


Fig. 6 EYR of fast growing willow crop used as a feed stock for electric production. At each step in the process an EYR can be calculated if all the inputs are known (dashed lines). The final EYR (at the right) is the ratio that the user receives.

usually not accounted for by other energy evaluation methods. Nonemergy approaches most often evaluate only nonrenewable resources and often do not account for the free services that a system receives from the environment (e.g., the photosynthetic activity driven by the solar radiation or the dilution of pollutants by the wind), which are just as much a requirement for the productive process as are fossil fuels. Finally, most nonemergy methods do not have an accounting procedure for human

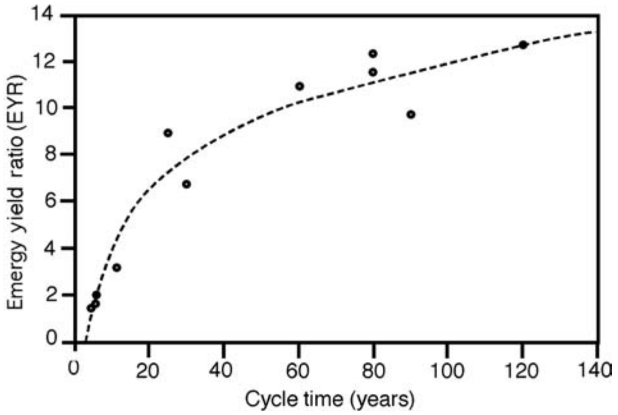


Fig. 7 The EYR is related to the cycle time of the resource. The graph shows the emergy yield ratio for 11 different forest cropping systems having very different cycle times, from willow systems of a couple of years to rainforest tress that require well over 100 years to mature. As the cycle time gets longer, the quality of the wood increases and the EYR increases.

labor, societal services, and information. Emergy accounting includes all of these, perhaps not perfectly, but it places them in perspective and, thus, helps us understand the huge network of supporting energies necessary to support any particular economic activity in our culture.

The idea that a calorie of sunlight is not equivalent to a calorie of fossil fuel or electricity, or even to a calorie of human work, strikes many people as preposterous, because they believe that a calorie is a calorie is a calorie. Some have rejected the concept as being impracticable, because from their perspective, it is impossible to quantify the amount of sunlight that is required to produce a quantity of oil. Others reject it because emergy does not appear to conform to first-law accounting principles. The concept of energy quality has been most controversial, and emergy and transformity have been even more so. Although quality has been recognized somewhat in the energy literature, in which different forms of fossil energy are expressed in coal or oil equivalents (and some researchers have even expressed electricity in oil equivalents by using first-law efficiencies), there has been widespread rejection of quality corrections of other forms of energy.

The emergy approach represents a conceptual framework that is absolutely needed for a reliable investigation of the interplay of natural ecosystems and human-dominated systems and processes. The common thread is the ability to evaluate all forms of energy, materials, and human services on a common basis by converting them to equivalents of one form of energy: solar emergy, a measure of the past and present environmental support to

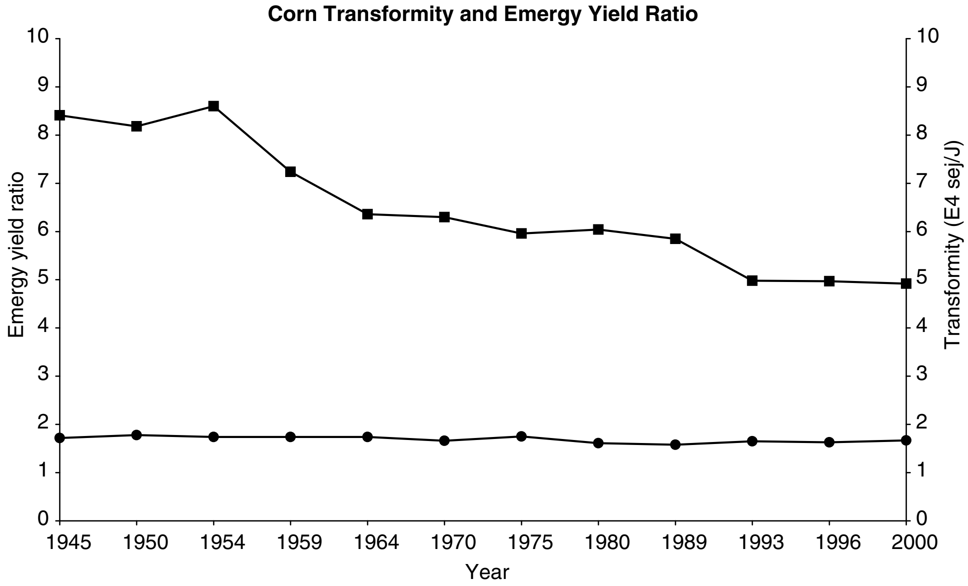


Fig. 8 Historical perspective on transformity and EYR of corn grown in the U.S.A. The transformity has declined over the years reflecting increased efficiencies and yields per acre. However, the EYR has remained essentially static during this time at about 1.7–1.

any process occurring in the biosphere. Through this quality correction, it is possible to evaluate all the inputs to processes and compute true net yields for processes, including potential energy sources. Without quality correction, net energy accounting can evaluate fossil energy return only for fossil energy invested; it cannot include human services, materials, and environmental services, in essence accounting for only a portion of the required inputs. The result can easily be a false assumption of the contributions from energy sources, but more important, this reasoning could lead to the wasteful use of energies in a futile pursuit of Maxwell's Demon.

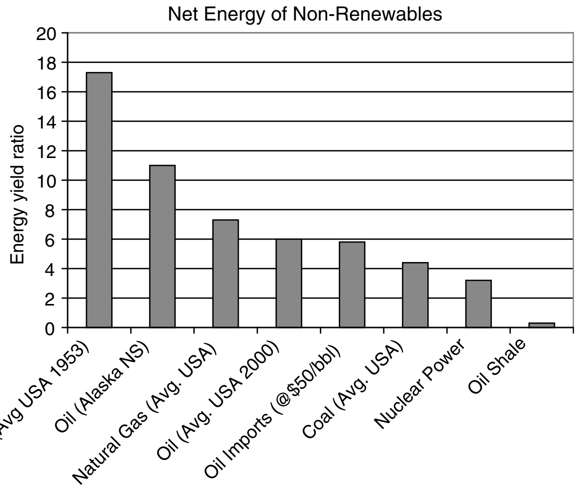


Fig. 9 Emergy yield ratios for major nonrenewable energy sources.

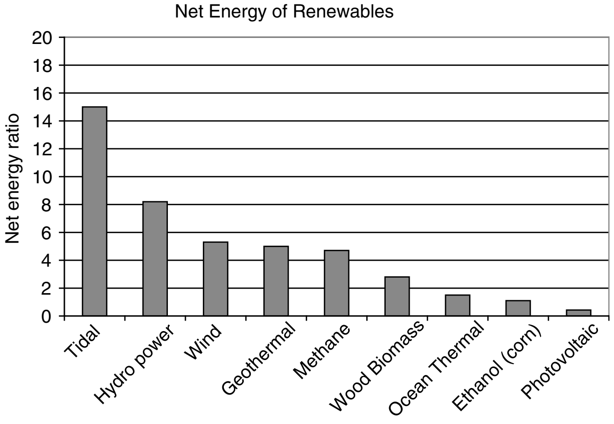


Fig. 10 Emergy yield ratios for renewable energy sources.

REFERENCES

1. Odum, H.T. *Environment, Power and Society*; Wiley: New York, 1971; 336.
2. Odum, H.T. Maximum power and efficiency: a rebuttal. Ecol. Model. **1983**, *20*, 71–82.
3. Odum, H.T. *Ecological and General Systems: An Introduction to Systems Ecology*; University Press of Colorado: Niwot, Colorado, 1994; [644, Revised edition of Systems Ecology, 1983, Wiley].
4. Odum, H.T. *Environmental Accounting. Emergy and Environmental Decision Making*; Wiley: New York, NY, 1996.
5. Odum, H.T.; Odum, E.C. *A Prosperous Way Down: Principles and Policies*; University Press of Colorado: Niwot, Colorado, 2001.

6. Brown, M.T.; Ulgiati, S. Emergy evaluation and environmental loading of electricity production systems. J. Clean. Prod. **2002**, *10*, 321–334.
7. Odum, H.T.; Brown, M.T.; Williams, S.B. *Handbook of Emergy Evaluation: A Compendium of Data for Emergy Computation Issued in a Series of Folios. Folio *1—Introduction and Global Budget*; Center for Environmental Policy, Environmental Engineering Sciences, University of Florida: Gainesville, FL, 2000; 16.
8. Odum, H.T. *Handbook of Emergy Evaluation: A Compendium of Data for Emergy Computation Issued in a Series of Folios. Folio #2—Emergy of Global Processes.* Center for Environmental Policy, Environmental Engineering Sciences, University of Florida: Gainesville, FL, 2000. British Petroleum *BP Statistical Review of World Energy*; The British Petroleum Company: London, 1997; 41.
9. Odum, H.T. Energy, ecology and economics. AMBIO **1973**, *2* (6), 220–227.
10. Doherty, S.J. Emergy Evaluations of And Limits to Forest Production. Ph.D. Dissertation, Environmental Engineering Sciences, University of Florida, Gainesville, 1995; 215.
11. Pimentel, D.; Hurd, L.E.; Bellotti, A.C.; et al. Food production and the energy crisis. Science **1973**, *182*, 443–449.
12. Pimentel, D.; Pimentel, M. *Food, Energy and Society*, Wiley: New York, 1979; 165.
13. Pimentel, D.; Warnaeke, A.F.; Teel, W.S.; et al. Food versus biomass fuel: socioeconomic and environmental impacts in the United States, Brazil, India and Kenya. Adv. Food Res. **1988**, *32*, 185–238.
14. Pimentel, D.; Wen, D. Technological changes in energy use in U.S. agricultural production. In *Agroecology*; Carrol, C.R., Vandermeer, J.H., Rosset, P.M., Eds., McGraw Hill: New York, NY, 1990; 147–164.
15. Giampietro, M.; Ulgiati, S.; Pimentel, D. Feasibility of large-scale biofuel production. Does an enlargement of scale change the picture? BioScience **1997**, *47* (9), 587–600.
16. Sclater, J.F.; Taupart, G.; Galson, I.D. The heat flow through the oceanic and continental crust and the heat loss of the earth. Rev. Geophys. Space Phys. **1980**, *18*, 269–311.
17. Miller, G.A. The flux of tidal energy out of the deep oceans. J. Geophys. Res. **1966**, *71*, 2485–2489 [Gainesville, 30].
18. Campbell, D. Emergy analysis of human carrying capacity and regional sustainability: an example using the state of Maine (Appendix). Environ. Monit. Assess. **1998**, *51*, 531–569.
19. Wiin-Nielsen, A.; Chen, T. *Fundamentals of Atmospheric Energetics*, Oxford Press: New York, 1993; 376.
20. Ryabchikov, A. *The Changing Face of the Earth*, Progress Publishers: Moscow, 1975; 203; [Translated by J. Williams].
21. Kinsman, B. *Wind, Waves, Their Generation and Propagation on the Ocean Surface*; Prentice-Hall: Englewood Cliffs, NJ, 1965; 676.
22. Brown, M.T.; McClanahan, T. Emergy analysis perspectives for Thailand and Mekong river dam proposals. Ecol. Model. **1996**, *91*, 105–130.
23. Brown, L.R.; Renner, M.; Flavin, C. *Vital Signs, 1997. The Environmental Trends That Are Shaping Our Future*; W.W. Morton & Co: New York, 1997; 165.
24. Leith, H.; Whittaker, R.H. *Primary Productivity of the Biosphere*; Springer: New York, 1975.
25. Oldeman, L.R. The global extent of soil degradation. In *Soil Resilience and Sustainable Land Use*; Greenland, D.J., Szabolcs, I., Eds., CAB International: Wallington, U.K., 1994; 561.
26. Mannion, A.M. *Agriculture and Environmental Change: Temporal and Spatial Dimensions*; Wiley: New York, 1995; 405.
27. USDI. *Mineral Commodity Summaries, January 1997*; U.S. Department of Interior: Washington, DC, 1996.
28. World Resources Institute *World Resources 1996–97*; Oxford University Press: New York, 1996.
29. Odum, E.C.; Odum, H.T. System of ethanol production from sugarcane in Brazil. Ciencia e Cultura **1984**, *37* (11), 1849–1855.
30. Brown, M.T.; Ulgiat, S. Emergy evaluation of the biosphere and natural capital. Ambio **1999**, *28* (6), 486–493.

BIBLIOGRAPHY

1. Oort, A. H.; Ascher, S. C.; Levitus, S.; Peixoto, J. P. New estimates of the available potential energy in the World Ocean. J. Geophys. Res. **1989**, *94*, 3187–3200.
2. Todd, D. K.; ed. The Water Encyclopedia, 2nd Ed.; Lewis Publication: Chelsea, MI, 1970.

Emissions Trading

Paul M. Sotkiewicz
Public Utility Research Center, University of Florida, Warrington College of Business, Gainesville, Florida, U.S.A.

Abstract
Emission trading is a system of rights or permits that gives the holder the right to emit 1 unit of a designated pollutant. Permits or rights to pollute then can be considered an input to production and are priced like any other commodity. The idea behind emissions trading is to meet environmental goals at the lowest possible cost compared with other environmental policies, such as command-and-control (CAC) or emissions taxes. Simple in concept, emissions trading can become complex in practice, considering all elements that must be in place. The most widely used trading mechanism is cap-and-trade, wherein the overall emissions level is capped and permits can be traded between emissions sources. Other forms of trading, though not used as often as cap-and-trade, include offset or project trading and emission rate trading. Early experience with emissions trading via offset or project-based trading, as well as the current experience with cap-and-trade, have resulted in significant cost savings to participants vs traditional CAC methods but have not achieved all the cost savings available.

INTRODUCTION: IDEAS AND CONTEXT BEHIND EMISSIONS TRADING

The idea of emissions trading, popularized by Dales[1] and then formalized by Montgomery,[2] is to create a system of property rights or permits—or, as they are called in many trading programs, allowances—in the spirit of Coase[3] that would give the holders of the rights/permits the right to emit 1 unit of a pollutant. These rights/permits/allowances can be thought of as inputs to production much like any other input, such as coal, oil, or natural gas, and thus would have a market-determined price and be tradable like any other commodity. These rights have value because the number of rights available is limited (capped) either explicitly or implicitly. As shown by Montgomery[2] and reproduced in Baumol and Oates,[4] emissions trading has the property of meeting an aggregate emissions (reduction) target at the lowest possible cost, because trading provides the ultimate flexibility to polluting sources in how best to meet the emissions target. Sources not only have flexibility in choosing technologies or input mixes to minimize the cost of meeting emission targets at individual sources, but also can buy and sell the permits/rights/allowances to pollute among one another to allocate the burden of emissions reductions in such a way as minimize the cost in aggregate across all sources. A cost-minimizing allocation of emissions reductions results in sources with low costs of abatement making greater reductions and those with higher costs of abatement making fewer reductions than they may otherwise make under command-and-control (CAC) policies. Thus, it can be said that the cost-saving benefits of emission trading vis-à-vis CAC policies is greater in proportion with the variability in emissions control costs.

As a policy option to achieve environmental compliance with pollution reduction goals, emissions trading is relatively new in its widespread application, though the first trading programs go back to the mid-1970s and have been used in a variety of contexts.[5] Prior to the launching of the first emissions trading schemes, the policy option to meet environmental objectives came in the form of CAC regulations that required emissions sources to meet a legislated emissions rate standards or to meet a stated technology standard. On a larger scale, the early policy for air pollution in the United States, beginning in 1970, mandated that specified concentration levels of pollutants be attained and then maintained at or below those levels going forward under the National Ambient Air Quality Standards (NAAQS). Many areas were in nonattainment of the standards, which would not permit the entry of new emission sources that would be associated with economic growth.[6] Consequently, the first emissions trading scheme, an offset policy or emission reduction credit (ERC) trading mechanism, was born out of the necessity to accommodate economic growth while moving toward attainment of the NAAQS in the middle 1970s.[7] The system was quite simple in concept. Existing sources in an area could reduce their emissions below an administratively defined baseline level and then could sell those offsets or ERCs to a new source entering the area at a price agreed upon by the parties. A variant of offset trading known as a bubble was introduced in 1979. The bubble provided flexibility to allocate emissions among multiple sources at the same facility (e.g., multiple generating units

Keywords: Emissions trading; Cap-and-trade; Emissions reduction credits; Offset trading; Least-cost emissions reduction.

at the same plant), so long as total facility emissions did not exceed a specified level.[8]

The movement to emissions trading as a policy option has also been driven by the cost of CAC policies relative to the least-cost way of meeting emissions standards. As shown in Portney,[6] a multitude of studies conducted during the 1980s showed ratios of CAC cost to least cost in a range from as low as 1.07 to as high as 22. The movement toward widespread application and acceptance of cap-and-trade programs led by the Title IV Sulfur Dioxide (SO_2) Trading Program (SO_2 Program) from the 1990 Clean Air Act Amendments (CAAA) can be seen as the meeting of environmental interests that want to see further emissions reductions with business and political interests that want to see market-driven policies.[9]

TYPES OF TRADING MECHANISMS

Cap-and-Trade

Under a cap-and-trade scheme, the aggregate level of emissions is capped, and property rights/permits/allowances are created such that the number of allowances available does not exceed the cap. Examples of cap-and-trade markets include the markets facilitated by the U.S. Environmental Protection Agency (USEPA), including the current SO_2 Program and NO_x SIP Call Program and the soon-to-be implemented Clean Air Interstate Rule (CAIR) and Clean Air Mercury Rule trading programs,[10] the Regional Clean Air Incentives Market (RECLAIM) in California,[11] and the European Union's Emissions Trading Scheme (EU ETS).[12] Cap-and-trade programs are perhaps the most used and visible of all emissions trading programs.

Offset or Project-Based Trading

In an offset or project-based trading scheme similar to that described above, potential emissions sources create credits by reducing emissions below their administratively determined baselines, so that credits can be sold to other sources that may be emitting more than their baselines. The ERC generated in this scheme is generally not a uniform commodity like the permit/property right/allowance that is defined under a cap-and-trade regime, but the number of ERCs created or needed is often determined on a project (case-by-case) basis. The spirit of an offset scheme is to cap emissions implicitly, though this is likely not the case in practice.[8] In-depth descriptions of such programs for the United States can be found in Hahn and Hester[13] and Environmental Law Institute.[14] An example in the context of carbon policy is the Clean Development Mechanism (CDM).[15]

Emissions Rate-Based Trading

In a rate-based trading environment, an emissions rate standard (e.g., lbs/mmBtu) is determined that must be met in aggregate, but sources can create credits by reducing emissions rates below the standard and sell these to sources with emissions rates above the standard. An example of this type of trading program exists for electric utility nitrogen oxide (NO_x) sources subject to Title IV of the 1990 CAAA.[16] Under this program, sources within the same company may trade credits to meet the NO_x emissions rate standard. Because credits are being traded to meet the standard, emission are in general not capped.[8]

ELEMENTS OF EMISSIONS TRADING PROGRAMS

As cap-and-trade emissions trading programs are the most prevalent, active, and visible, most of the elements in trading regimes are described with cap-and-trade in mind, though many of these elements also relate to other forms of trading in many cases. The format of this section closely follows U.S. EPA.[8]

Definition of Affected Sources

Determination of the emission sources to be included in the program (affected sources) is essential. Ideally, as many emissions sources as possible should be included in any trading program, but consideration must also be given to the size of the source, ability to monitor and report emissions from the source, and any other considerations that may be deemed important. Under the SO_2 Program, for example, existing simple cycle combustion turbines and steam units less than 25 MW in capacity were exempt from the program. One could surmise that such technologies were not large sources of SO_2 emissions or were too small to monitor in a cost-effective manner.

Measurement, Verification, and Emissions Inventory

Without the ability to measure emissions, emissions trading programs would not be workable. The measurement of emissions for the inventory can be done through a monitoring system or through the use of mass-balance equations. To verify emission monitoring, results can be checked against mass-balance equation derived emissions readings to ensure robust readings. The measurement of emissions prior to the commencement of a trading program can help provide a basis by which to set a cap and allocate permits/allowances in a cap-and-trade system, to set a baseline by which the emissions reductions can be measured in an offset system, or to determine emissions rates.

Determination of an Emissions Cap

In cap-and-trade systems, the element that makes emissions reductions valuable is the programwide limit on total emissions. The decision on the level of the emissions cap is as much political as it is scientific. In an ideal world with perfect information, the cap would be set so that the net benefits to society would be maximized (marginal costs of emissions reductions would equal the marginal benefits of reduction). Determining benefits is not as easy as determining costs of pollution reduction, however, though great strides have been made in recent years. As a matter of practice, although consideration is given to maximizing net benefits to society, the level of the cap is often determined through political means to gain wider stakeholder acceptance.[9]

Unit of Trade: Allowance/Permit/Emissions Reduction Credit

To facilitate trading among sources, it is crucial to define the units of trade between emissions sources. In the academic literature, these units are sometimes called permits. In the language of the U.S. EPA, these units are known as allowances in cap-and-trade systems, and as ERCs in offset and bubble systems in the United States. Regardless of the nomenclature, a permit/allowance/ERC gives the holder the right to emit 1 unit of pollutant where units can be defined in pounds, tons, kilograms, or any other accepted unit of measure. In effect, the allowance/permit/ERC is a property right to pollute and can be traded between sources at a price amenable to the parties, as any other commodity could be.

Compliance Period and True-Up

The time period for which emissions are to be controlled must be defined. For emissions in the SO_2 Trading Program, the compliance period is January 1–December 31, whereas in the NO_x OTC Market, the predecessor to the current NO_x SIP Call Market, it was May 1–September 31.[16] Sources must have allowances at least equal to their emissions during the compliance period. A trading regime may also allow a true-up period, during which sources may verify their actual emissions during the compliance period and then buy or sell allowances for the purposes of meeting the just-concluded compliance-period obligations.

Allowance/Permit Allocation or ERC Baseline

Under cap-and-trade, permits/allowances must be allocated to affected sources, or in the case of an offset system, the baseline must be established by which reductions are measured and ERCs are created.

With respect to cap-and-trade, there are three primary allocation methods: historical baseline, fixed; auction; and historical baseline with updating. Allocations may also be created for new units, or as a reward for undertaking certain actions to reduce emissions quickly or by other means. Under historical baseline, fixed methods, the allocation is *gratis* and is determined by a measure of performance for affected sources from the past. The performance measure could be based on output or input. Being based on the past, affected sources cannot engage in any behavior in an attempt to gain larger allowance allocations. For Phase I units in the SO_2 Program announced in 1990, for example, allocations were based on an emissions rate per unit of heat input from 1985 to 1987.

Under an auction allocation method, the allowances are sold directly to sources at a predetermined interval in advance of the time when affected sources will need the allowances to cover their emissions.

Under an updating methodology, allowance allocations beyond the first years of the program are determined based on updated performance measures such as heat input or output, rather than being permanently fixed to historic performance. Some countries in the EU ETS, for example, have decided to use an updating allocation method in which sources that are shut down permanently will have their allocations taken away.[17]

Choosing a baseline is crucial for offset programs, as the baseline determines how many ERCs are created through abatement. The determination of what the baseline might be varies across jurisdictions and is often open to negotiation in U.S.-based programs.[14]

Spatial and Temporal Trading Rules

The wider trading opportunities across space and time are, the greater is the potential for cost savings from trading. Still, political or environmental considerations may necessitate rules defining and restricting how trade can be made across space and time. If the pollutant being traded is seen to create greater damages where it is concentrated (such as mercury) or may become highly concentrated due to wind and weather patterns (NO_x and SO_2), it may be necessary to create spatial trading ratios that differ from a one-to-one exchange or to restrict trades from one zone to another, as has been done in the RECLAIM program.[18]

The ability to create ERCs or to save allowances for future use is known as banking. Banking ERCs or allowances is a way of trading between time periods and is allowed in many programs. Such a practice is warranted if concentration increases at a given point in time are not troublesome. But if pollutant concentrations increase at a given point in time, such as NO_x during summer ozone

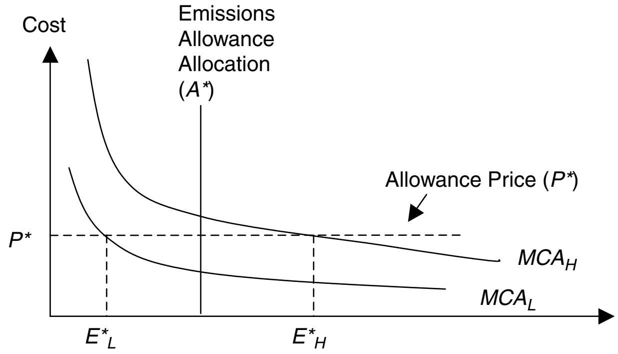


Fig. 1 Benefits from emissions trading.

season, banking may not be allowed such as in RECLAIM or by some states in the NO_x SIP Call Program.[18]

Penalties and Enforcement

All affected sources in a trading program must possess enough allowances to cover its emissions in a cap-and-trade program. Penalties and enforcement are necessary because without penalties or enforcement, there is no reason for sources to hold the necessary allowances to be in compliance. In cap-and-trade systems, a penalty per allowance not held, well in excess of the market price of allowances, for any shortfalls in allowances is necessary so that sources will participate in the market and maintain the emissions cap, and will not *de facto* opt out by paying no penalty or a small penalty.

FIRM INCENTIVES UNDER EMISSIONS TRADING

Consider a cap-and-trade system for electric generating units in which allowances have already been allocated, but keep in mind that the same logic applies to offset and emissions rate trading systems. If a generating unit has low abatement costs, that unit can reduce emissions below its allowance allocation and sell the remaining allowances or simply bank them for future use. As long as the marginal (incremental) cost of abatement (emissions reduction) is less than the allowance price, it pays the generating unit to reduce emissions further and sell the freed-up allowance, as shown in Fig. 1.

Marginal cost of abatement (MCA) MCA_L represents a low marginal abatement cost source. Being allocated $A*$ allowances, if the market price of allowances is $P*$, it pays a generating unit that has low abatement costs to reduce emissions until it reaches E_L^*. The revenue from allowance sales is the rectangle with the width $A^* - E_L^*$ and height $P*$. The cost to the utility company is the area under MCA_L between $A*$ and E_L^*. The net profit from the allowance sale is the area of the revenue rectangle above MCA_L. Conversely, a unit may have high abatement costs, represented by MCA_H. Rather than reduce pollution, that unit may find it less expensive to buy allowances in the open market and use the purchased allowances, along with the allowance allocation, to cover its emissions obligation. Units will continue buying allowances as long as the marginal (incremental) cost of abatement (emissions reduction) is greater than the allowance price. A more formal way of expressing this idea is that the unit with high abatement costs (Fig. 1) will buy $E_H^* - A^*$ allowances in the market at the price $P*$. That unit's expenditure on allowances is the rectangle with width $E_H^* - A^*$ and height $P*$. Because of its reduction in abatement costs, the area between $A*$ and E_H^* and below MCA_H, is greater than the expenditure on allowances, and the unit with high abatement costs will benefit. Also note that the allowance market leads to the equalization of the marginal costs of abatement across generating units.

COST-MINIMIZING POLLUTION ABATEMENT WITH EMISSIONS TRADING

Consider the following example of two firms with the objective of minimizing the cost of achieving the aggregate emissions restriction of 2000 tons. (Table 1). Let E_i in Table 1 represent the unrestricted or baseline emissions level for firm i. Let e_i be the emissions level of firm i after abatement, so that abatement for firm i is equal to $(E_i - e_i)$.

The least-cost solution for emissions abatement can be solved by minimizing the cost of abatement subject to the aggregate emissions restriction:

$$\mathrm{Min}_{e_1,e_2} C_1(E_1 - e_1) + C_2(E_2 - e_2)$$

$$\text{s.t. } e_1 + e_2 \leq A^*$$

The solution to this problem requires that the MCA be equalized across the firms, as shown in the solution to this problem in Table 2. Also note in Table 2 that Firm 2 makes much larger reductions (2000 vs 400) than Firm 1, as its cost of abatement is only a fifth of that for Firm 1.

Em—Energy Con

Table 1 Two-firm cost-minimizing example

	Firm 1		Firm 2
Unrestricted/baseline emissions (tons) E_i	2000		2400
Total cost of abatement function	$C_1(E_1-e_1)=0.5(E_1-e_1)^2$		$C_2(E_2-e_2)=0.1(E_2-e_2)^2$
Marginal cost of abatement function	$\text{MCA}_1=(E_1-e_1)$		$\text{MCA}_2=0.2(E_2-e_2)$
Aggregate emission restriction		$e_1+e_2\leq 2000$	

Now consider a cap-and-trade emissions trading program. Let X_i be the allowance allocation for firm i and x_i be the allowance purchase ($x_i>0$) or allowance sales ($x_i<0$) position of firm i. Let P be the price of allowances in the market. Each firm in the market minimizes its cost of pollution abatement and allowance purchases/sales subject to the restriction that emissions, e_i, are less than or equal to the allowance allocation plus the net position:

$$\text{Min}_{e_i,x_i} C_i(E_i - e_i) + Px_i$$

$$\text{s.t. } e_i \leq X_i + x_i$$

The solution to this problem for each firm requires that its MCA be equal to the allowance price P just as shown in Fig. 1, where the allowance price is the mechanism by which marginal costs of abatement are equalized across firms. Additionally, the aggregate emissions constraint must be satisfied $\Sigma_i\, e_i \leq \Sigma_i X_i$, and assuming no banking, the sum of allowance sales and purchases is equal to zero $\Sigma_i\, x_i=0$.

Extending the example in Table 1, assume that each firm is initially allocated 1000 allowances signifying the right to emit 1000 ton. We know that each firm reduces emissions up to the point where MCA=P, and the MCAs are equal across firms. Consequently, we arrive at the same emissions outcome and MCA as the least-cost solution in Table 2. This results in Firm 2's having 600 surplus allowances, which it sells to Firm 1, which needs 600 allowances at a price of 400/ton (MCA). Table 3 shows the result.

It is important to note that the allowance purchases and sales cancel each other out in aggregate and that the actual abatement cost is the same as the least-solution found in Table 2.

An important lesson from the results in Table 3 is that emissions trading can achieve the least-cost solution without the need to collect detailed information on sources' abatement costs, and as we will see below, the method by which allowances are allocated does not change this result.

Allowance Allocation and Distribution of Costs

How allowances are allocated across firms, whether they are allocated gratis or by auction, the distribution of the initial allocation, once determined, does not change the aggregate abatement cost, although updating methods introduce other inefficiencies and effects, as discussed in Ahman et al.[17] and Burtraw et al.[19] Shifting allocations does change the distribution of the cost burden to meet the aggregate emissions constraint, however. In the previous example, we assumed that each firm was allocated 1000 allowances. Suppose instead that Firm 1 is allocated all 2000 allowances, and Firm 2 gets none. This does not change the optimizing behavior on how much is emitted; neither does it change the aggregate abatement cost. What is does do is change the allowance position of each of the firms: Firm 1 sells 400 ton, giving it allowance revenue of 160,000, and Firm 2 buys 400 ton, adding 160,000 in allowance costs. The allowance price, P, remains unchanged at 400. All that has changed is the distribution of the cost burden in meeting the aggregate emissions constraint.

Suppose that instead of there being a *gratis* allocation of allowances, the allowances were auctioned off, and the government kept the revenue for use elsewhere, such as offsetting other taxes. In this case, the allocations X_1 and X_2 are equal to zero, and the net allowance position for each firm is equal to the number of allowances they would need to satisfy their emissions constraints. Once again, the change in allocation method does not change the

Table 2 Least-cost solution to the two-firm example

	Firm 1		Firm 2
Emissions level, e_i	1,600		400
Abatement level, (E_i-e_i)	400		2,000
Total cost of abatement	80,000		400,000
Marginal cost of abatement	400		400
Aggregate abatement cost		480,000	

Table 3 Solution to the two-firm emission-trading example

	Firm 1		Firm 2
Allowance allocation, X_i	1,000		1,000
Emissions level, e_i	1,600		400
Abatement level, $(E_i - e_i)$	400		2,000
Allowance position, x_i	600		−600
Total cost of abatement	80,000		400,000
Marginal cost of abatement	400		400
Allowance price		400	
Allowance costs	240,000		−240,000
Aggregate abatement cost		480,000	

optimizing behavior of firms, as shown in Table 3, as they still produce the same emissions ($e_1 = 1600$, $e_2 = 400$); neither does it change the allowance price, which is still $P = 400$. What does change is the allowance cost for the firms. Under the gratis allocation firms are, in effect, being allocated a subsidy in the sense of not needing to pay for any costs associated with their emissions covered by the allocation as they would under an auction scheme. Under an auction, firms pay the government directly for their emissions through the purchase of allowances at auction. Given the optimal emissions levels and the allowance price, Firm 1 would pay 640,000 in allowance costs at auction, and Firm 2 would pay 160,000 in allowance costs, providing the government 800,000 in auction proceeds that were forgone with the gratis allocation scheme.

Emissions Trading Vs CAC

Suppose that the environmental regulator promulgated a CAC regime in which each firm had to reduce its emissions by 1200 tons, and an equal share of the reductions needed to get emissions down to 2000 tons. Such a regime leads to certainty regarding the emissions level, but mandating each firm to reduce by the same amount (in total quantity or percentage terms) is quite unlikely to lead to the least-cost solution. Table 4 shows the results of the above CAC scheme.

The aggregate abatement cost under this CAC regime is almost double the cost from emissions trading (864,000 vs 480,000). The MCAs are not equalized under CAC; the MCAs indicate that Firm 2 should engage in more abatement and Firm 1 should engage in less abatement activity in an effort to equalize the marginal costs across firms.

The only way in which the CAC regime could achieve the least-cost solution is to collect detailed information on the costs of abatement at the firm or source level so as to implement the least-cost outcome as the CAC target.

Emissions Trading Vs Emissions Taxes

Rather than using CAC or emissions trading to reduce emissions, the environmental authority wants to employ emissions taxes to reduce emissions. The incentives under emissions taxes are similar to those under emission trading, as shown in Fig. 1. Firms will want to reduce emissions until the marginal cost of abatement is equal to the tax rather than the allowance price. The difference between the two regimes involves the certainty with which an emissions target will be met. Under emissions trading, there is certainty about the emissions resulting from the program, assuming no banking, but the allowance price is uncertain, as it is determined endogenously. With emissions taxes, the price of emissions is certain, but the resulting emission level is determined endogenously.

Table 4 Command-and-control (CAC) costs

	Firm 1		Firm 2
Required reductions $(E_i - e_i)$	1,200		1,200
Emissions level, e_i	800		1,200
Total cost of abatement	720,000		144,000
Marginal cost of abatement	1,200		240
Aggregate abatement cost		864,000	

Table 5 Emissions tax of 300/ton results

	Firm 1		Firm 2
Emissions level, e_i	1,700		900
Reductions $(E_i - e_i)$	300		1,500
Total cost of abatement	45,000		81,000
Marginal cost of abatement	300		300
Aggregate emissions		2,600	
Aggregate abatement cost		126,000	

Suppose that the environmental regulator imposes an emission tax of 300 per ton. By design, the marginal costs of abatement are equalized across firms, thus minimizing the cost of meeting the uncertain emissions level. Table 5 shows the result for the tax of 300 per ton.

The resulting emissions of 2600 are greater than the target set forth under either emission trading or CAC, although this higher emissions level is achieved at least-cost. If the goal is to achieve the 2000-ton limit with emissions taxes, this would require a constant adjustment of the tax level until the goal is met. Such adjustments to the tax, however, would introduce uncertainty and increase risk for firms operating in their respective industries, and would likely be fought by the owners of the affected sources.

EXPERIENCE WITH EMISSIONS TRADING PROGRAMS AND CONCLUDING THOUGHTS

The early experiences with offset trading programs were that the programs achieved cost savings, but many opportunities for cost savings went unexploited due to administrative complexity and burden, and that the environmental improvements were not as great as was hoped.[7,13] More recent programs have seen little trading, as other environmental programs have resulted in greater reductions, reducing the demand for credits.[14]

Burtraw et al.[18] provides a survey of the performance of U.S. cap-and-trade programs. Ellerman et al.[9] offers a comprehensive analysis of the early years of the SO_2 Program. Burtraw and Evans[16] offer insight into federal NO_x trading programs. Overall, there is general agreement that the cap-and-trade programs in the United States have offered significant cost savings and technological innovation, and have resulted in significant emissions reductions. Moreover, no emissions hot spots or locally high concentrations have been found, as were feared by environmentalists. Still, there is a growing consensus that the existing programs have not achieved all the possible cost savings from trading, One possible explanation offered by Sotkiewicz and Holt[20] is that affected sources also face economic regulation by state public utility commissions which may provide incentives to affected sources that lead to deviations from the overall cost minimizing solution.

The EU ETS was only 18 months into operation at the time this article was written, and little can be said about its performance or the performance of the CDM to date. Still, the movement of the EU toward emissions trading, based on the U.S. experience, shows confidence in emissions trading; it also shows that the experience to date has been more positive than negative, and has delivered reduced emissions at lower cost than traditional CAC regimes.

REFERENCES

1. Dales, J.H. *Pollution, Property, and Prices*; University of Toronto Press: Toronto, Ont., Canada, 1968.
2. Montgomery, W.D. Markets in licenses and efficient pollution control programs. J. Econ. Theory **1972**, *5*, 395–418.
3. Coase, R.H. The problem of social cost. J. Law Econ. **1960**, *3*, 1–44.
4. Baumol, W.J.; Oates, W.E. *The Theory of Environmental Policy*, 2nd Ed.; Cambridge University Press: Cambridge, U.K., 1988.
5. Stavins, R.N. Market-based environmental policies. In *Public Policies for Environmental Protection*; 2nd Ed; Portney, P.R., Stavins, R.N., Eds.; Resources for the Future: Washington, DC, 2000.
6. Portney, P.R. Air pollution policy. In *Public Policies for Environmental Protection*; 2nd Ed; Portney, P.R., Stavins, R.N., Eds.; Resources for the Future: Washington, DC, 2000.
7. Hahn, R.W. Economic prescriptions for environmental problems: how the patient followed the doctor's orders. J. Econ. Perspect. **1988**, *3* (2), 95–114.
8. U.S. Environmental Protection Agency (USEPA). Tools of the trade: a guide to designing and operating a cap and trade program for pollution control, EPA430-B-03-002; Available at: www.epa.gov/airmarkets/international/tools.pdf (accessed on November 2006).
9. Ellerman, A.; et al. *Markets for Clean Air: The U.S. Acid Rain Program*; Cambridge University Press: Cambridge, U.K., 2000.
10. U.S. Environmental Protection Agency (USEPA). Clean Air Markets Program home page. Available at: www.epa.gov/airmarkets (accessed on November 2006).

Em–Energy Con

11. South Coast Air Quality Management District (SCAQMD). RECLAIM home page. Available at: www.aqmd.gov/reclaim/index.htm (accessed on November 2006).
12. European Commission. Emission Trading Scheme (EU ETS) home page. Available at: http://ec.europa.eu/environment/climat/emission.htm (accessed on November 2006).
13. Hahn, R.W.; Hester, G.L. Where did all the markets go? An analysis of EPA's emissions trading program. Yale J. Regul. **1989**, *6* (109), 109–153.
14. Environmental Law Institute (ELI). *Emissions Reduction Credit Trading Systems: An Overview of Recent Results and an Assessment of Best Practices*, Washington, DC, (accessed on November 2006).
15. CDMWatch. *The Clean Development Mechanism (CDM) Toolkit: A Resource for Stakeholders, Activists, and NGOs*. Available at: www.cdmwatch.org/files/CDMToolkitVO19-02-04.pdf.
16. Burtraw, D.; Evans, D.A. NO_x emissions in the United States. In *Choosing Environmental Policy: Comparing Instruments and Outcomes in the United States and Europe*; Harrington, W., Morgenstern, R.D., Sterner, T., Eds.; Resources for the Future: Washington, DC, 2004.
17. Ahman, M. et al. *The Ten Year Rule: Allocation of Emission Allowances in the EU Emission Trading System*, Resources for the Future Discussion Paper 05–30, Washington, DC, June 2005.
18. Burtraw, D. et al. *Economics of Pollution Trading for SO_2 and NO_x*, Resources for the Future Discussion Paper 05–05, Washington, DC, January 2005.
19. Burtraw, D. et al. *The Effect of Allowance Allocation on the Cost of Carbon Emission Trading*, Resources for the Future Discussion Paper 01–30, Washington, DC, April 2001.
20. Sotkiewicz, P.M.; Holt, L. Public utility commission regulation and cost effectiveness of title IV: lessons for CAIR. Electricity J. **2005**, *18* (8), 68–80.

Energy Codes and Standards: Facilities

Rosemarie Bartlett
Pacific Northwest National Laboratory, Richland, Washington, U.S.A.
Mark A. Halverson
Pacific Northwest National Laboratory, West Halifax, Vermont, U.S.A.
Diana L. Shankle
Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.

Abstract
Energy codes and standards play a vital role in the marketplace by setting minimum requirements for energy-efficient design and construction. They outline uniform requirements for new buildings as well as additions and renovations. This article covers basic knowledge of codes and standards; development processes of each; adoption, implementation, and enforcement of energy codes and standards; and voluntary energy efficiency programs.

INTRODUCTION

Energy-efficient buildings offer energy, economic, and environmental benefits. They reduce energy expenditures and environmental pollutants. They also create economic opportunities for business and industry by promoting new energy-efficient technologies.

Unfortunately, the marketplace does not guarantee energy-efficient design and construction. Owners of commercial buildings generally pass on energy costs to consumers or tenants, eliminating any incentive for energy-efficient design and construction. Homebuyers often are motivated more by up-front costs than operating costs.

Energy codes and standards play a vital role by setting minimum requirements for energy-efficient design and construction. They outline uniform requirements for new buildings as well as additions and renovations.

THE DIFFERENCE BETWEEN ENERGY CODES AND ENERGY STANDARDS, AND THE MODEL ENERGY CODE

Energy codes—specify how buildings must be constructed or perform and are written in mandatory, enforceable language. States or local governments adopt and enforce energy codes for their jurisdictions. Residential and commercial energy codes typically include requirements for building envelopes, mechanical systems, service water heating, and lighting and electrical power.

Energy standards—describe how buildings *should* be constructed to save energy cost-effectively. They are published by national organizations such as the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE). They are not mandatory, but serve as national recommendations, with some variation for regional climate. States and local governments frequently use energy standards as the technical basis for developing their energy codes. Some energy standards are written in mandatory, enforceable language, making it easy for jurisdictions to incorporate the provisions of the energy standards directly into their laws or regulations. Residential and commercial energy standards typically include requirements for building envelopes, mechanical systems, service water heating, and lighting and electrical power.

The model energy code (MEC)[*]—The International Code Council (ICC) publishes and maintains the International Energy Conservation Code (IECC), which is an MEC that makes allowances for different climate zones. Because it is written in mandatory, enforceable language, state and local jurisdictions can easily adopt the model as their energy code. Before adopting the IECC, state and local governments often make changes to reflect regional building practices.

Table 1 provides an overview of energy standards and the MEC.

How are Energy Standards Developed and Revised?

Standards 90.1 and 90.2 are developed and revised through voluntary consensus and public hearing processes that are critical to widespread support for their adoption.

Keywords: Energy; Code; Standard; ASHRAE; IECC; ICC; Adoption; Enforcement.

* MEC in this article refers to any model energy code, not specifically to the predecessor to the IECC.

Table 1 Overview of national energy standards and the model energy code (MEC)

Title	Type	Sponsoring organization(s)	Description	Commonly used versions
International Energy Conservation Code (IECC)	MEC	International Code Council (ICC)	Applies to residential and commercial buildings. Written in mandatory, enforceable language	1998 IECC 2000 IECC 2003 IECC 2006 IECC
American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE)/IESNA/ANSI standard 90.1: Energy-Efficient Design of New Buildings Except Low-Rise Residential Buildings	Energy standard	ASHRAE, together with the Illuminating Engineering Society of North America (IESNA) and the American National Standards Institute (ANSI)	Applies to all buildings except residential buildings with three stories or less	90.1-1989 90.1-1999[a] 90.1-2001 90.1-2004
ASHRAE Standard 90.2 Energy-Efficient Design of New Low-Rise Residential Buildings	Energy standard	ASHRAE	Applies to residential buildings with three stories or less	90.2-1993 90.2-2001 90.2-2004

[a]This and subsequent versions written in mandatory, enforceable language.

Who is Involved?

ASHRAE works with other standards organizations, such as the Illuminating Engineering Society of North America (IESNA), American National Standards Institute (ANSI), American Society of Testing and Materials (ASTM), Air Conditioning and Refrigeration Institute (ARI), and Underwriters Laboratories (UL). The voluntary consensus process also includes representation from other groups:

- The design community, including architects, lighting, and mechanical designers
- Members of the enforcement community, including building code officials, representatives of code organizations, and state regulatory agencies
- Building owners and operators
- Industry and manufacturers
- Utility companies
- Representatives from the U.S. Department of Energy (DOE), energy advocacy groups, and the academic community.

DOE's Role

Federal law requires the DOE to determine whether revisions to the residential portion of the IECC would improve energy efficiency in the nation's residential buildings and whether revisions to ASHRAE/IESNA/ANSI Standard 90.1 would improve energy efficiency in the nation's commercial buildings.

When DOE determines that a revision would improve energy efficiency, each state has 2 yr to review the energy provisions of its residential or commercial building code. For residential buildings, a state has the option of revising its residential code to meet or exceed the residential portion of the IECC. For commercial buildings, a state is required to update its commercial code to meet or exceed the provisions of Standard 90.1.

How does the Process Work?

Standards 90.1 and 90.2 are both on continuous maintenance and are maintained by separate Standing Standards Project Committees. Committee membership varies from 10 to 60 voting members. Committee membership includes representatives from the list above to ensure balance among all interest categories.

After the committee proposes revisions to the standard, the revised version undergoes public review and comment. The committee usually incorporates nonsubstantive changes into the standard without another review. Substantive changes require additional public review. Occasionally, mediation is necessary to resolve differing views.

When a majority of the parties substantially agree (known as consensus), the revised standard is submitted for approval to the ASHRAE board of directors. Those not in agreement with the decision may appeal to the board. If an appeal is upheld, further revision, public comment, and resolution occur. If the board denies the appeal, publication of the revised standard proceeds.

What's the Timing of Revisions to Standards 90.1 and 90.2?

Standards 90.1 and 90.2 are automatically revised and published every 3 yr. However, anyone may propose a revision at any time. Approved interim revisions (called addenda) are posted on the ASHRAE Web site and are included in the next published version.

Key activities relating to revisions, including responding to public comments, typically occur during one of ASHRAE's annual (June) or midwinter (January) meetings. Public review of standards commonly occurs 2–4 months after one of these meetings.

HOW ARE MODEL ENERGY CODES DEVELOPED AND REVISED?

The most recent MECs are the 2003 IECC and the 2006 IECC. These are developed and published by the ICC through an open public-hearing process. Prior to 1998, the IECC was known as the Council of American Building Officials MEC.

Who is Involved?

The IECC Code Development Committee typically comprises 7–11 code, building science, and energy experts appointed by the ICC. Most, but not all, committee members are code officials. They may or may not be members of the ICC. The International Residential Code (IRC) Building and Energy Committee is approximately the same size, and includes builders, code officials, and industry representatives.

How does the Process Work?

Anyone may suggest a revision to the IECC or IRC by requesting a code change proposal from the committee and preparing a recommended change and substantiation. The committee publishes proposed changes and distributes them for review. This occurs about 6 weeks prior to an open public hearing, which is held in front of the code development committee.

At the public hearing, the committee receives testimony and then votes to approve, deny, or revise each change. The committee publishes its results.

Those wishing to have a proposed change reconsidered may submit a challenge to the committee's recommended action. Proponents and opponents present additional information at a second public hearing, followed by a vote by the full ICC membership. This outcome may be appealed to the ICC board of directors.

What's the Timing of the Process?

The IECC and IRC are revised on and 18-months cycle. However, full publication of the documents occurs every third yr, with supplements issued in the interim years. When developing and adopting their own energy codes, states and local governments typically adopt the fully published IECC or IRC. By specifically adopting the supplements as well, state and local governments ensure that their energy codes include important additions and clarifications to the IECC or IRC.

ADOPTION OF ENERGY CODES ON THE STATE AND LOCAL LEVEL

Before adopting or revising an energy code, states and local governments often assemble an advisory board comprising representatives of the design, building construction, and enforcement communities. This body determines which (if any) energy standards and MECs should be adopted. The group also considers the need to modify energy standards and MECs to account for local preferences and construction practices. The body also may serve as a source of information during the adoption process.

Overview of the Adoption Process

The adoption process generally includes the following steps:

- Change is initiated by a legislative or regulatory agency with authority to promulgate energy codes. Interested or affected parties also may initiate change. An advisory body typically is convened. The proposed energy code is developed.
- The proposal undergoes a legislative or public review process. Public review options include publishing a notice in key publications, filing notices of intent, and holding public hearings. Interested and affected parties are invited to submit written or oral comments.
- The results of the review process are incorporated into the proposal, and the final legislation or regulation is prepared for approval.
- The approving authority reviews the legislation or regulation. Revisions may be submitted to the designated authority for final approval or for filing.
- After being filed or approved, the code is put into effect, usually on some specified future date. This grace period allows those regulated to become familiar with any new requirements. The period between adoption and effective date typically varies from 30 days to 6 months.
- Details of the adoption process vary depending on whether the energy code is adopted by legislation, regulation, or a local government. Each is discussed below.

Adoption Through Legislation

State legislation rarely includes the complete text of an energy standard or MEC. More commonly, legislation references an energy standard or MEC that is already published. The legislation often adds administrative provisions addressing enforcement, updating, variances, and authority.

Another common approach is to use legislation to delegate authority to an agency, council, or committee. The delegated authority is empowered to develop and adopt regulations governing energy-related aspects of building design and construction. Such regulations are discussed in "Adoption through Regulation" later in this entry. Some states adopt the administrative provisions of the energy code by legislation and the technical provisions by regulation, or vice versa.

Adoption Through Regulation

A key factor in a state's ability to regulate the energy-related aspects of design and construction is the extent to which the state has authority over adoption, administration, implementation, and enforcement of building construction regulations. In most states, a single state agency has such authority. In some states, no such authority exists. If multiple state agencies, committees, or councils are involved, the authority is diluted.

When a state agency, council, or committee has authority to adopt regulations, it must follow requirements outlined in the legislation that enables development, revisions, and adoption of the regulations. The technical provisions of the regulation may be unique to the state, or the regulations may adopt, by reference, national energy standards or an MEC. When a state adopts regulations, it typically includes its own administrative provisions within the regulations.

Adoption by Local Government

If a state has limited authority to adopt an energy code [a "home rule" state (In the energy codes and standards arena, home rule means the state cannot interfere or control on the local level.)], units of local government have the option to assume that responsibility. Local governments also can adopt standards or codes that are more stringent than the state's.

A local government's municipal code typically includes a title or provision covering building construction, under which energy provisions can be adopted.

Most local governments adopt an MEC by reference. They apply administrative provisions from other building construction regulations to implementation and enforcement of the energy code.

Timing of the Adoption and Revision of State and Local Codes

Most states adopt or revise energy codes in concert with the publication of a new edition of a national energy standard or MEC. This may occur either through a regulatory process or automatically because state regulation or legislation refers to "the most recent edition."

Adoption also can be tied to the publication date of an energy standard or MEC; e.g., "This regulation shall take effect 1 month from publication of the adopted MEC."

IMPLEMENTATION OF ENERGY CODES ON THE STATE AND LOCAL LEVEL

During implementation, the adopting jurisdiction(s) must prepare building officials to enforce the energy code and prepare the building construction community to comply with it. It is important for all stakeholders to know that a new code is coming and understand what is required. Many states or jurisdictions start this education process several years in advance of an energy code change—often before adoption itself. The more publicity about and training on the new code there are, the more it will be accepted and used.

Communication and information exchange should occur in several contexts:

- Between the code-adopting bodies and the code-enforcing bodies
- Between the code-adopting bodies and the building construction community
- Between the code-enforcing bodies and the building construction community
- Within the building construction community and the code-enforcing bodies

Training is critical. To be effective, training must cater to the specific needs of building officials, architects, designers, engineers, manufacturers, builders and contractors, and building owners. Training for specific stakeholders can be provided or sponsored by the following:

- State energy offices and agencies
- Universities and community colleges
- Professional organizations and societies
- Utilities
- Trade associations
- National or regional code organizations
- Others, such as product distributors.

The DOE, the ICC, ASHRAE, and other codes organizations can supply tools and materials to make implementation and training easier for states and local jurisdictions.

ENFORCEMENT OF ENERGY CODES ON THE STATE AND LOCAL LEVELS

Enforcement ensures compliance with an energy code and is critical to securing energy savings. Enforcement strategies vary according to a state or local government's regulatory authority, resources, and manpower. Enforcement can include all or some of the following activities:

- Plan review
- Product, material, and equipment specifications review
- Testing and certification review
- Supporting calculation review
- Building inspection during construction
- Evaluation of materials substituted in the field
- Building inspection immediately prior to occupancy.

Sometimes a state or local government has no enforcement authority. The courts address enforcement if and when legal action is sought by a building owner against a designer or contractor.

State Enforcement

State enforcement is a common approach in smaller states, in rural jurisdictions that have no code officials, and for state-owned or financed construction. Enforcement by a single state agency usually is more uniform than enforcement conducted by several local agencies. Plan review is generally performed by one office. Although there may be numerous state field inspectors, they are bound under one organization. This arrangement benefits the building construction community by offering a single point of contact. However, if state resources are limited, plan reviews and construction inspections may not be performed as thoroughly as warranted.

Local Enforcement

Local enforcement agencies are closer to the construction site and in more direct contact with the design and construction community. This offers the potential for more regular enforcement during design and construction. However, local jurisdictions may lack sufficient resources to support enforcement. Because jurisdictions vary, local enforcement may lead to some noncompliance across a state. Compliance is enhanced when a state code agency actively supports local governments in their efforts to enforce the state code.

Some states allow local jurisdictions to petition to conduct enforcement activities that are usually the responsibility of the state. This strategy offers the advantages associated with state enforcement, recognizes those local governments with equivalent enforcement capabilities, and helps ensure comparable levels of compliance. Continued state oversight is necessary to ensure a consistent level of enforcement by local jurisdictions. A hybrid approach might involve the state conducting the plan review and the local authority conducting the construction inspection.

Third-Party Alternatives

Some states and local governments allow qualified third parties to conduct plan reviews. Often, these reviewers have more experience dealing with the complexities and subtleties of energy codes and standards; have better sources, references, and contacts because of affiliations with professional organizations; and can help ease heavy workloads.

VOLUNTARY ENERGY-EFFICIENCY PROGRAMS

Voluntary programs encourage a level of energy efficiency above code. They can help motivate consumers to recognize the value of energy efficiency. Examples include the following:

- Home energy rating systems—Also known as HERS, these compare the energy efficiency of a home with that of a computer-simulated reference house. The rating involves analysis of the home's construction plans and at least one onsite inspection. This information is used to estimate the home's annual energy costs and give the home a rating between 0 and 100. The lower the score, the more efficient the home.
- ENERGY STAR—The U.S. Environmental Protection Agency outlines criteria for ENERGY STAR certification of homes and commercial buildings. ENERGY STAR homes are typically 30% more energy efficient than average minimum energy codes. For more information, go to www.energystar.gov.
- Utility, government, and other programs—Utilities, state and local governments, and other organizations often sponsor programs that qualify buildings based on certain standards. Examples include the following:
  - The DOE's Building America Program, www.eere.energy.gov/buildings/building_america
  - The U.S. Green Building Council's Leadership in Energy and Environmental Design (LEED™), www.usgbc.org/DisplayPage.aspx?CategoryID=19
  - The New Buildings Institute Advanced Buildings Benchmark™ www.poweryourdesign.com/benchmark.htm
  - ASHRAE's Advanced Energy Design Guides, http://resourcecenter.ashrae.org/store/ashrae.

DOE SUPPORT

The DOE's Building Energy Codes Program (BECP) supports state and local governments in their efforts to implement and enforce building energy codes. This support includes the following activities:

- Developing and distributing easy-to-use compliance tools and materials
- Providing financial and technical assistance to help adopt, implement, and enforce building energy codes
- Participating in the development of MECs and energy standards
- Providing information on compliance products and training, and energy-code-related news.

For more information on BECP products and services, visit the BECP Web site at www.energycodes.gov.

CONCLUSION

Energy codes and standards set baseline requirements for energy-efficient design and construction. Several organizations play a role in their development, and there is a long, intertwined development history. The development processes vary, but suggested code and standard changes may be submitted by any interested party. Development is only the first step, however; adoption and enforcement are critical next steps.

BIBLIOGRAPHY

1. American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE), www.ashrae.org.
2. Association of Energy Engineers, www.aeecenter.org.
3. Bartlett, R.; Halverson, M.A.; Shankle, D.L. *Understanding Building Energy Codes and Standards*; Pacific Northwest National Laboratory: Richland, WA, 2003; [PNNL-14235].
4. Bartlett, R.; Halverson, M.A. *Codes 101*; Pacific Northwest National Laboratory: Richland, WA, 2006; [PNNL-SA-51223].
5. Codes 101. Available at: http://energycode.pnl.gov/moodle/course/view.php?id=7.
6. International Code Council, www.iccsafe.org.
7. National Conference of States on Building Codes and Standards (NCSBCS), www.ncsbcs.org.
8. Residential Energy Services Network, www.natresnet.org.
9. U.S. Department of Energy's Building Energy Codes Program, www.energycodes.gov.

Energy Conservation

Ibrahim Dincer
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology (UOIT), Oshawa, Ontario, Canada

Adnan Midilli*
Department of Mechanical Engineering, Faculty of Engineering, Nigde University, Nigde, Turkey

Abstract
This study highlights these issues and potential solutions to the current environmental issues; identifies the main steps for implementing energy conservation programs and the main barriers to such implementations; and provides assessments for energy conservation potentials for countries, as well as various practical and environmental aspects of energy conservation.

INTRODUCTION

Civilization began when people found out how to use fire extensively. They burned wood and obtained sufficiently high temperatures for melting metals, extracting chemicals, and converting heat into mechanical power, as well as for cooking and heating. During burning, the carbon in wood combines with O_2 to form carbon dioxide (CO_2), which then is absorbed by plants and converted back to carbon for use as a fuel again. Because wood was unable to meet the fuel demand, the Industrial Revolution began with the use of fossil fuels (e.g., oil, coal, and gas). Using such fuels has increased the CO_2 concentration in the air, leading to the beginning of global warming. Despite several warnings in the past about the risks of greenhouse-gas emissions, significant actions to reduce environmental pollution were not taken, and now many researchers have concluded that global warming is occurring. During the past two decades, the public has became more aware, and researchers and policymakers have focused on this and related issues by considering energy, the environment, and sustainable development.

Energy is considered to be a key catalyst in the generation of wealth and also a significant component in social, industrial, technological, economic, and sustainable development. This makes energy resources and their use extremely significant for every country. In fact, abundant and affordable energy is one of the great boons of modern industrial civilization and the basis of our living standard. It makes people's lives brighter, safer, more comfortable, and more mobile, depending on their energy demand and consumption. In recent years, however, energy use and associated greenhouse-gas emissions and their potential effects on the global climate change have been of worldwide concern.

Problems with energy utilization are related not only to global warming, but also to such environmental concerns as air pollution, acid rain, and stratospheric ozone depletion. These issues must be taken into consideration simultaneously if humanity is to achieve a bright energy future with minimal environmental impact. Because all energy resources lead to some environmental impact, it is reasonable to suggest that some (not all) of these concerns can be overcome in part through energy conservation efforts.

Energy conservation is a key element of energy policy and appears to be one of the most effective ways to improve end-use energy efficiency, and to reduce energy consumption and greenhouse-gas emissions in various sectors (industrial, residential, transportation, etc.). This is why many countries have recently started developing aggressive energy conservation programs to reduce the energy intensity of the their infrastructures, make businesses more competitive, and allow consumers to save money and to live more comfortably. In general, energy conservation programs aim to reduce the need for new generation or transmission capacity, to save energy, and to improve the environment. Furthermore, energy conservation is vital for sustainable development and should be implemented by all possible means, despite the fact that it has its own limitations. This is required not only for us, but for the next generation as well.

Considering these important contributions, the energy conservation phenomenon should be discussed in a comprehensive perspective. Therefore, the main objective of this article is to present and discuss the world's primary energy consumption and production; major environmental problems; potential solutions to these issues;

Keywords: Energy; Energy conservation; Environment; Sustainability; Policies; Strategies; Life-cycle costing.
* Currently on sabbatical leave at University of Ontario Institute of Technology (UOIT).

practical energy conservation aspects; research and development (R&D) in energy conservation, energy conservation, and sustainable development; energy conservation implementation plans; energy conservation measurements; and life-cycle costing (LCC) as an excellent tool in energy conservation. In this regard, this contribution aims to:

- Help explain main concepts and issues about energy conservation
- Develop relations between energy conservation and sustainability
- Encourage energy conservation strategies and policies
- Provide energy conservation methodologies
- Discuss relations between energy conservation and environmental impact
- Present some illustrative examples to state the importance of energy conservation and its practical benefits

In summary, this book contribution highlights the current environmental issues and potential solutions to these issues; identifies the main steps for implementing energy conservation programs and the main barriers to such implementations; and provides assessments for energy conservation potentials for countries, as well as various practical and environmental aspects of energy conservation.

WORLD ENERGY RESOURCES: PRODUCTION AND CONSUMPTION

World energy consumption and production are very important for energy conservation in the future. Economic activity and investment patterns in the global energy sector are still centered on fossil fuels, and fossil-fuel industries and energy-intensive industries generally have been skeptical about warnings of global warming and, in particular, about policies to combat it. The increase of energy consumption and energy demand indicates our dependence on fossil fuels. If the increase of fossil-fuel utilization continues in this manner, it is likely that the world will be affected by many problems due to fossil fuels. It follows from basic scientific laws that increasing amounts of CO_2 and other greenhouse gases will affect the global climate. The informed debate is not about the existence of such effects, but about their magnitudes and seriousness. At present, the concentration of CO_2 is approximately 30% higher than its preindustrial level, and scientists have already been able to observe a discernible human influence on the global climate.[1]

In the past, fossil fuels were a major alternative for overcoming world energy problems. Fossil fuels cannot continue indefinitely as the principal energy sources, however, due to the rapid increase of world energy demand and energy consumption. The utilization distribution of fossil-fuel types has changed significantly over the past 80 years. In 1925, 80% of the required energy was supplied from coal, whereas in the past few decades, 45% came from petroleum, 25% from natural gas, and 30% from coal. Due to world population growth and the advance of technologies that depend on fossil fuels, reserves of those fuels eventually will not be able to meet energy demand. Energy experts point out that reserves are less than 40 years for petroleum, 60 years for natural gas, and 250 years for coal.[2] Thus, fossil-fuel costs are likely to increase in the near future. This will allow the use of renewable energy sources such as solar, wind, and hydrogen. As an example, the actual data[3,4] and projections of world energy production and consumption from 1980 to 2030 are displayed in the following figures, and the curve equations for world energy production and consumption are derived as shown Table 1.

As presented in Figs. 1 and 2, and in Table 2, the quantities of world primary energy production and consumption are expected to reach 14,499.2 and 13,466.5 Mtoe, respectively, by 2030. World population is now over six billion, double that of 40 years ago, and it is likely to double again by the middle of the 21st century. The world's population is expected to rise to about seven billion by 2010. Even if birth rates fall so that the world population becomes stable by 2030, the population still would be about ten billion. The data presented in Figs. 1 and 2 are expected to cover current energy needs provided that the population remains constant. Because the population is expected to increase dramatically, however, conventional energy resource shortages are likely to occur, due to insufficient fossil-fuel resources. Therefore, energy

Table 1 World energy production and consumption models through statistical analysis

Energy	Production (Mtoe)	Correlation coefficient	Consumption (Mtoe)	Correlation coefficient
World primary	$=148.70\times Year-287{,}369$	0.998	$=139.62\times Year-269{,}953$	0.998
World oil	$=44.47\times Year-85{,}374$	0.997	$=42.18\times Year-80{,}840$	0.998
World coal	$=20.05\times Year-37{,}748$	0.946	$=23.18\times Year-43{,}973$	0.968
World NG	$=45.73\times Year-89{,}257$	0.999	$=46.27\times Year-90{,}347$	0.999

Mtoe, million tons of oil equivalent.

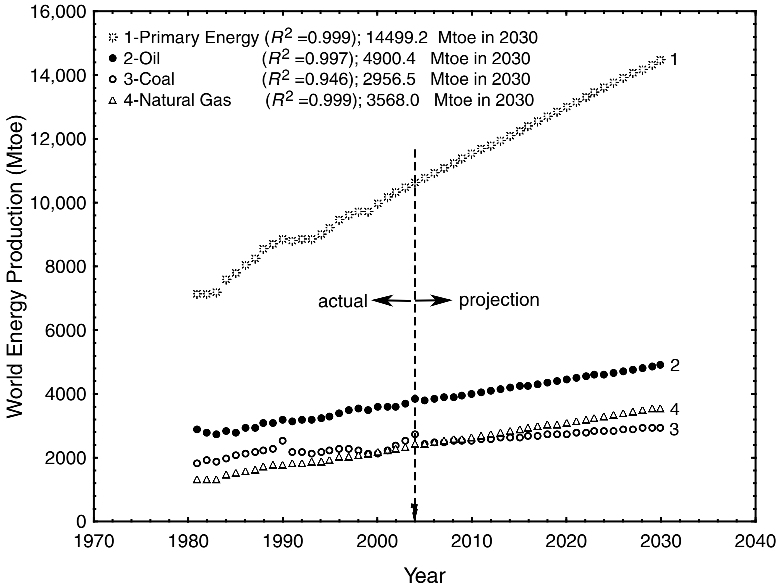


Fig. 1 Variation of actual data taken from Refs. 3,4, and projections of annual world energy production. Mtoe, million tons of oil equivalent.
Source: From Refs. 3,4.

conservation will become increasingly important to compensate for shortages of conventional resources.

MAJOR ENVIRONMENTAL PROBLEMS

One of the most important targets of modern industrial civilizations is to supply sustainable energy sources and to develop the basis of living standards based on these energy sources, as well as implementing energy conservation measures. In fact, affordable and abundant sustainable energy makes our lives brighter, safer, more comfortable, and more mobile because most industrialized and developing societies use various types of energy. Billions of people in undeveloped countries, however, still have limited access to energy. India's per-capita consumption of electricity, for example, is one-twentieth that of the United States. Hundreds of millions of Indians live "off the grid"-that is, without electricity-and cow dung is still a major fuel for household cooking. This continuing reliance on such preindustrial energy sources is also one of the major causes of environmental degradation.[5]

After many decades of using fossil fuels as a main energy source, significant environmental effects of fossil fuels became apparent. The essential pollutants were from greenhouse gases (e.g., CO_2, SO_2, and NO_2). Fossil fuels are used for many applications, including industry, residential, and commercial sectors. Increasing fossil-fuel utilization in transportation vehicles such as automobiles, ships, aircrafts, and spacecrafts has led to increasing pollution. Gas, particulate matter, and dust clouds in the atmosphere absorb a significant portion of the

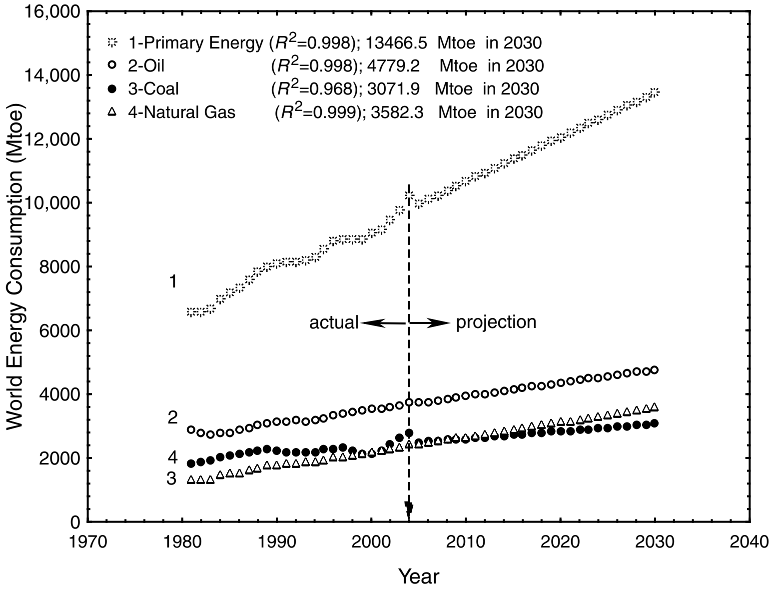


Fig. 2 Variation of actual data taken from Refs. 3,4, and projections of annual world energy consumption. Mtoe, million tons of oil equivalent.
Source: From Refs. 3,4.

Table 2 Some extracted values of world primary and fossil energy production and consumption

Year	Primary energy production (Mtoe)	Primary energy consumption (Mtoe)	Oil prod. (Mtoe)	Oil cons. (Mtoe)	Coal prod. (Mtoe)	Coal cons. (Mtoe)	NG prod. (Mtoe)	NG cons. (Mtoe)
1994	8,996.9	8,310.1	3237.1	3204.4	2178.1	2185.5	1891.2	1876.7
2000	9,981.9	9,079.8	3614.0	3538.7	2112.4	2148.1	2189.9	2194.5
2006	10,930.3	10,115.7	3833.1	3767.0	2475.3	2515.7	2470.5	2471.8
2012	11,822.5	10,953.4	4099.9	4020.0	2595.6	2654.8	2744.9	2749.5
2018	12,714.7	11,791.1	4366.7	4273.1	2715.9	2793.8	3019.2	3027.1
2024	13,606.9	12,128.8	4633.5	4526.1	2836.2	2932.9	3293.6	3304.7
2030	14,499.2	13,466.5	4900.4	4779.2	2956.5	3071.9	3568.0	3582.3

Mtoe, million tons of oil equivalent.

solar radiation directed at Earth and cause a decrease in the oxygen available for the living things. The threat of global warming has been attributed to fossil fuels.[2] In addition, the risk and reality of environmental degradation have become more apparent. Growing evidence of environmental problems is due to a combination of factors.

During the past two decades, environmental degradation has grown dramatically because of the sheer increase of world population, energy consumption, and industrial activities. Throughout the 1970s, most environmental analysis and legal control instruments concentrated on conventional pollutants such as SO_2, NO_x, particulates, and CO. Recently, environmental concern has extended to the control of micro or hazardous air pollutants, which are usually toxic chemical substances and harmful in small doses, as well as to that of globally significant pollutants such as CO_2. Aside from advances in environmental engineering science, developments in industrial processes and structures have led to new environmental problems.[6,7] In the energy sector, for example, major shifts to the road transport of industrial goods and to individual travel by cars has led to an increase in road traffic and, hence, to a shift in attention paid to the effects and sources of NO_x and to the emissions of volatile organic compounds (VOC). In fact, problems with energy supply and use are related not only to global warming, but also to such environmental concerns as air pollution, ozone depletion, forest destruction, and emission of radioactive substances. These issues must be taken into consideration simultaneously if humanity is to achieve a bright energy future with minimal environmental impact. Much evidence exists to suggest that the future will be negatively impacted if humans keep degrading the environment. Therefore, there is an intimate connection among energy conservation, the environment, and sustainable development. A society seeking sustainable development ideally must utilize only energy resources that cause no environmental impact (e.g., that release no emissions to the environment). Because all energy resources lead to some environmental impact, however, it is reasonable to suggest that some (not all) of the concerns regarding the limitations imposed on sustainable development by environmental emissions and their negative impacts can be overcome in part through energy conservation. A strong relation clearly exists between energy conservation and environmental impact, because for the same services or products, less resource utilization and pollution normally are associated with higher-efficiency processes.[8]

Table 3 summarizes the major environmental problems-such as acid rain, stratospheric ozone depletion, and global climate change (greenhouse effect)-and their main sources and effects.

As shown in Fig. 3, the world total CO_2 production is estimated to be 18,313.13 million tons in 1980, 25,586.7 million tons in 2006, 27,356.43 million tons in 2012, and 29,716.1 million tons in 2020 whereas fossil-fuel

Table 3 Major environmental issues and their consequences

Issues	Description	Main sources	Main effects
Acid precipitation	Transportation and deposition of acids produced by fossil-fuel combustion (e.g., industrial boilers, transportation vehicles) over great distances through the atmosphere via precipitation on the earth on ecosystems	Emissions of SO_2, NO_x, and volatile organic compounds (VOCs) (e.g., residential heating and industrial energy use account for 80% of SO_2 emissions)	Acidification of lakes, streams and ground waters, resulting in damage to fish and aquatic life; damage to forests and agricultural crops; and deterioration of materials, e.g., buildings, structures
Stratospheric ozone depletion	Distortion and regional depletion of stratospheric ozone layer though energy activities (e.g., refrigeration, fertilizers)	Emissions of CFCs, halons (chlorinated and brominated organic compounds) and N_2O (e.g., fossil fuel and biomass combustion account for 65%–75% of N_2O emissions)	Increased levels of damaging ultraviolet radiation reaching the ground, causing increased rates of skin cancer, eye damage and other harm to many biological species
Greenhouse effect	A rise in the earth's temperature as a result of the greenhouse gases	Emissions of carbon dioxide (CO_2), CH_4, CFCs, halons, N_2O, ozone and peroxyacetylnitrate (e.g., CO_2 releases from fossil fuel combustion (∼50% from CO_2), CH_4 emissions from increased human activity)	Increased the earth's surface temperature about 0.6°C over the last century and as a consequence risen sea level about 20 cm (in the next century by another 2°C–4°C and a rise between 30 and 60 cm); resulting in flooding of coastal settlements, a displacement of fertile zones for agriculture and food production toward higher latitudes, and a decreasing availability of fresh water for irrigation and other essential uses

Source: From Refs. 9–11.

consumption is found to be 6092.2 million tons in 1980, 8754.5 million tons in 2006, 9424.3 million tons in 2012, and 10,317.2 million tons in 2020. These values show that the CO_2 production will probably increase if we continue utilizing fossil fuel. Therefore, it is suggested that certain energy conversion strategies and technologies should be put into practice immediately to reduce future environmental problems.

The climate technology initiative (CTI) is a cooperative effort by 23 Organization for Economic Co-operation and

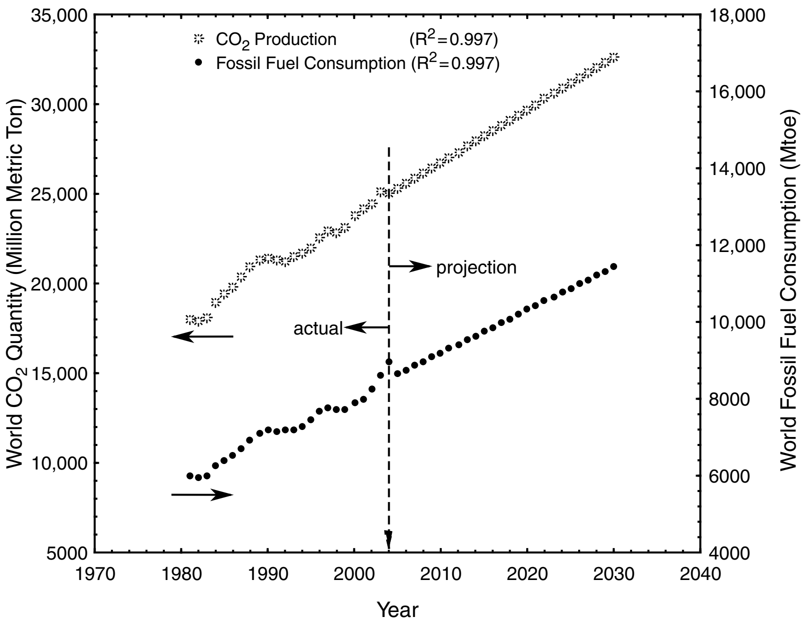


Fig. 3 Variation of world total fossil-fuel consumption and CO_2 production; actual data from Ref. 3 and projections. Mtoe, million tons of oil equivalent.
Source: From Ref. 3.

Development (OECD)/International Energy Agency (IEA) member countries and the European Commission to support the objectives of the united nations framework convention on climate change (UNFCCC). The CTI was launched at the 1995 Berlin Conference of the Parties to the UNFCCC. The CTI seeks to ensure that technologies to address climate change are available and can be deployed efficiently. The CTI includes activities directed at the achievement of seven broad objectives:

- To facilitate cooperative and voluntary actions among governments, quasigovernments, and private entities to help cost-effective technology diffusion and reduce the barriers to an enhanced use of climate-friendly technologies
- To promote the development of technology aspects of national plans and programs prepared under the UNFCCC
- To establish and strengthen the networks among renewable and energy efficiency centers in different regions
- To improve access to and enhance markets for emerging technologies
- To provide appropriate recognition of climate-friendly technologies through the creation of international technology awards
- To strengthen international collaboration on short-, medium-, and long-term research; development and demonstration; and systematic evaluation of technology options
- To assess the feasibility of developing longer-term technologies to capture, remove, or dispose of greenhouse gases; to produce hydrogen from fossil fuels; and to strengthen relevant basic and applied research

POTENTIAL SOLUTIONS TO ENVIRONMENTAL ISSUES

Although there are a large number of practical solutions to environmental problems, three potential solutions are given priority, as follows[11]:

- Energy conservation technologies (efficient energy utilization)
- Renewable energy technologies
- Cleaner technologies

In these technologies, we pay special attention to energy conservation technologies and their practical aspects and environmental impacts. Each of these technologies is of great importance, and requires careful treatment and program development. In this work, we deal with energy conservation technologies and strategies in depth. Considering the above priorities to environmental solutions, the important technologies shown in Fig. 4 should be put into practice.

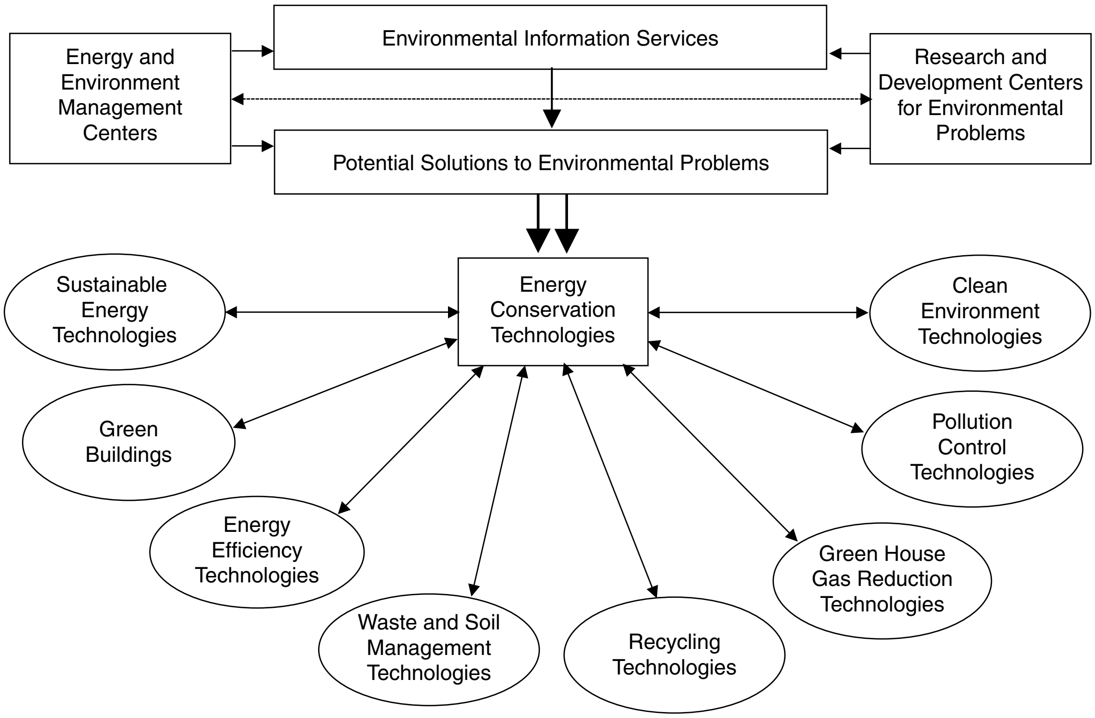


Fig. 4 Linkages between possible environmental and energy conservation technologies.

Em—Energy Con

PRACTICAL ENERGY CONSERVATION ASPECTS

The energy-saving result of efficiency improvements is often called energy conservation. The terms efficiency and conservation contrast with curtailment, which decreases output (e.g., turning down the thermostat) or services (e.g., driving less) to curb energy use. That is, energy curtailment occurs when saving energy causes a reduction in services or sacrifice of comfort. Curtailment is often employed as an emergency measure. Energy efficiency is increased when an energy conversion device-such as a household appliance, automobile engine, or steam turbine-undergoes a technical change that enables it to provide the same service (lighting, heating, motor drive, etc.) while using less energy. Energy efficiency is often viewed as a resource option like coal, oil, or natural gas. In contrast to supply options, however, the downward pressure on energy prices created by energy efficiency comes from demand reductions instead of increased supply. As a result, energy efficiency can reduce resource use and environmental impacts.[12]

The quality of a country's energy supply and demand systems is increasingly evaluated today in terms of its environmental sustainability. Fossil-fuel resources will not last indefinitely, and the most convenient, versatile, and inexpensive of them have substantially been used up. The future role of nuclear energy is uncertain, and global environmental concerns call for immediate action. OECD countries account for almost 50% of total world energy consumption: Current use of oil per person averages 4.5 bbl a year worldwide, ranging from 24 bbl in the United States and 12 bbl in western Europe to less than 1 bbl in sub-Saharan Africa. More than 80% of worldwide CO_2 emissions originate in the OECD area. It is clear, then, that OECD countries should play a crucial role in indicating a sustainable pattern and in implementing innovative strategies.[11]

From an economic as well as an environmental perspective, energy conservation holds even greater promise than renewable energy, at least in the near-term future. Energy conservation is indisputably beneficial to the environment, as a unit of energy not consumed equates to a unit of resources saved and a unit of pollution not generated.

Furthermore, some technical limitations on energy conservation are associated with the laws of physics and thermodynamics. Other technical limitations are imposed by practical technical constraints related to the real-world devices that are used. The minimum amount of fuel theoretically needed to produce a specified quantity of electricity, for example, could be determined by considering a Carnot (ideal) heat engine. However, more than this theoretical minimum fuel may be needed due to practical technical matters such as the maximum temperatures and pressures that structures and materials in the power plant can withstand.

As environmental concerns such as pollution, ozone depletion, and global climate change became major issues in the 1980s, interest developed in the link between energy utilization and the environment. Since then, there has been increasing attention to this linkage. Many scientists and engineers suggest that the impact of energy-resource utilization on the environment is best addressed by considering exergy. The exergy of a quantity of energy or a substance is a measure of the usefulness or quality of the energy or substance, or a measure of its potential to cause change. Exergy appears to be an effective measure of the potential of a substance to impact the environment. In practice, the authors feel that a thorough understanding of exergy and of how exergy analysis can provide insights into the efficiency and performance of energy systems is required for the engineer or scientist working in the area of energy systems and the environment.[8] Considering the above explanations, the general aspects of energy conservation can be summarized as shown in Fig. 5.

RESEARCH AND DEVELOPMENT STATUS ON ENERGY CONSERVATION

Now we look at R&D expenditures in energy conservation to assess the importance attached to energy conservation in the long range. The share of energy R&D expenditures going into energy conservation, for example, has grown greatly since 1976, from 5.1% in 1976 to 40.1% in 1990 and 68.5% in 2002.[11] This indicates that within energy R&D, research on energy conservation is increasing in importance. When R&D expenditures on energy conservation are compared with expenditures for research leading to protection of the environment in the 2000s, the largest share was spent on environment research. In fact, it is not easy to interpret the current trends in R&D expenditures, because energy conservation is now part of every discipline from engineering to economics. A marked trend has been observed since the mid-1970s, in that expenditures for energy conservation research have grown significantly, both in absolute terms and as a share of total energy R&D. These expenditures also grew more rapidly than those for environmental protection research, surpassing it in the early 1980s. Therefore, if R&D expenditures reflect long-term concern, there seems to be relatively more importance attached to energy conservation as compared with environmental protection.

In addition to the general trends discussed above, consider the industrial sector and how it has tackled energy conservation.

The private sector clearly has an important role to play in providing finance that could be used for energy efficiency investments. In fact, governments can adjust

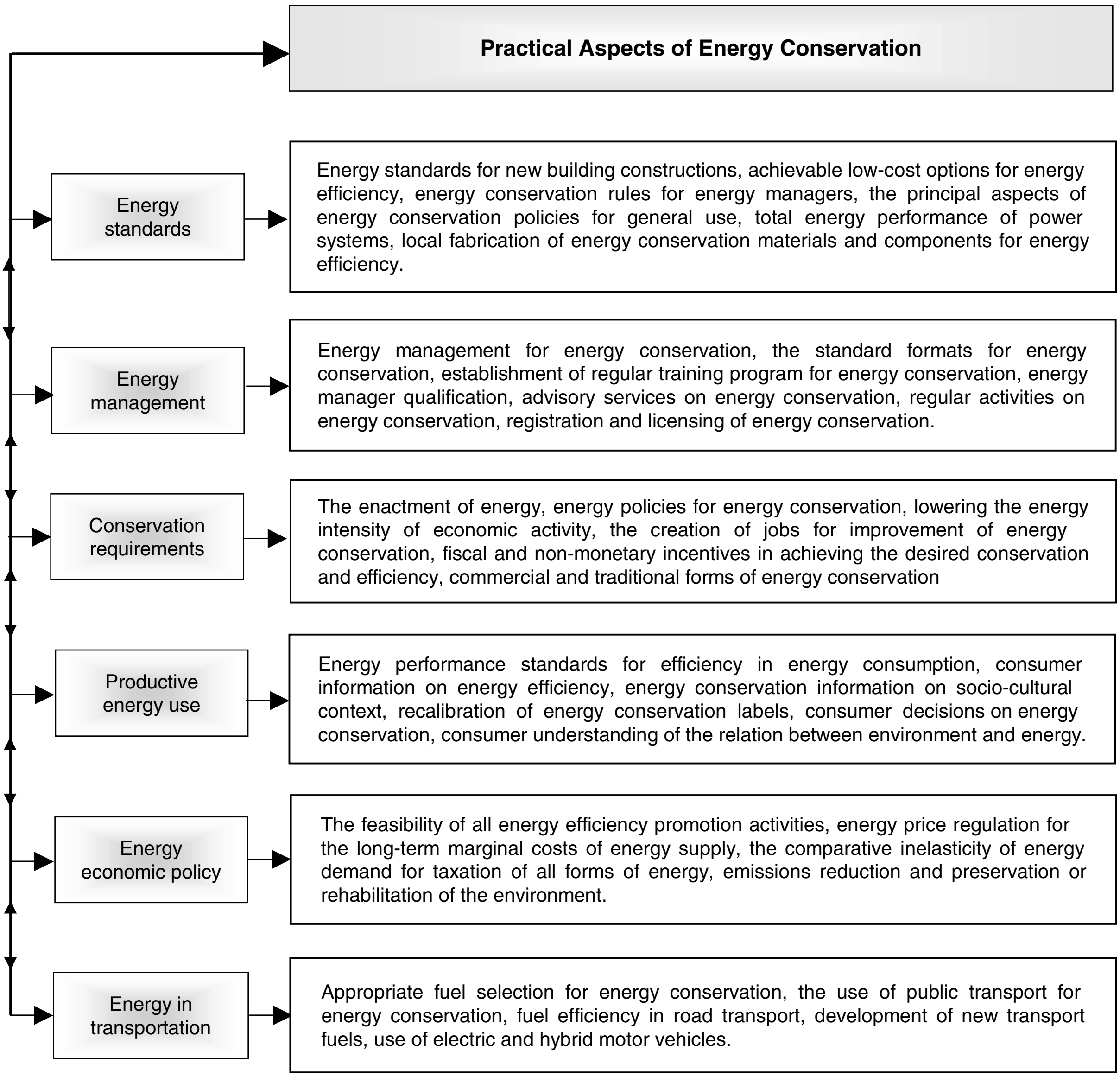


Fig. 5 A flow chart of practical energy conservation aspects.

their spending priorities in aid plans and through official support provided to their exporters, but they can influence the vast potential pool of private-sector finance only indirectly. Many of the most important measures to attract foreign investors include reforming macroeconomic policy frameworks, energy market structures and pricing, and banking; creating debt recovery programs; strengthening the commercial and legal framework for investment; and setting up judicial institutions and enforcement mechanisms. These are difficult tasks that often involve lengthy political processes.

Thus, the following important factors, which are adopted from a literature work[13] can contribute to improving energy conservation in real life. Fig. 6 presents the improvement factors of energy conservation.

ENERGY CONSERVATION AND SUSTAINABLE DEVELOPMENT

Energy conservation is vital for sustainable development and should be implemented by all possible means, despite the fact that it has its own limitations. This is required not only for us, but for the next generation as well.

A secure supply of energy resources is generally considered a necessity but not a sufficient requirement for development within a society. Furthermore, sustainable development demands a sustainable supply of energy resources that, in the long term, is readily and sustainably available at reasonable cost and can be utilized for all required tasks without causing negative societal impact. Supplies of such energy resources as fossil fuels (coal, oil,

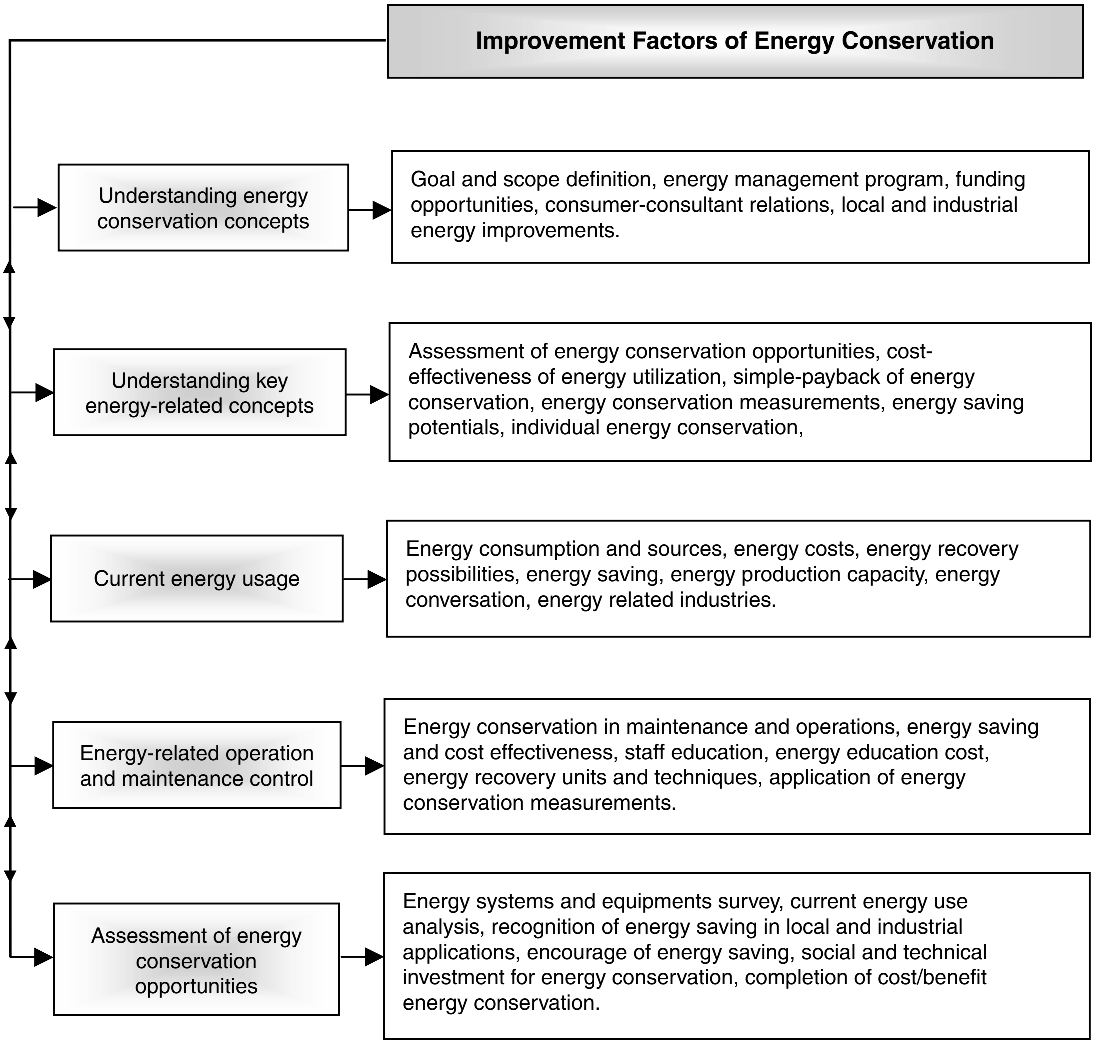


Fig. 6 Improvement factors of energy conservation.

and natural gas) and uranium are generally acknowledged to be finite. Other energy sources (such as sunlight, wind, and falling water) are generally considered to be renewable and, therefore, sustainable over the relatively long term. Wastes (convertible to useful energy forms through, for example, waste-to-energy incineration facilities) and biomass fuels usually also are viewed as being sustainable energy sources. In general, the implications of these statements are numerous and depend on how the term *sustainable* is defined.[14]

Energy resources and their utilization are intimately related to sustainable development. For societies to attain or try to attain sustainable development, much effort must be devoted not only to discovering sustainable energy resources, but also to increasing the energy efficiencies of processes utilizing these resources. Under these circumstances, increasing the efficiency of energy-utilizing devices is important. Due to increased awareness of the benefits of efficiency improvements, many institutes and agencies have started working along these lines. Many energy conservation and efficiency improvement programs have been developed and are being developed to reduce present levels of energy consumption. To implement these programs in a beneficial manner, an understanding is required of the patterns of "energy carrier" consumption-for example, the type of energy carrier used, factors that influence consumption, and types of end uses.[15]

Environmental concerns are an important factor in sustainable development. For a variety of reasons, activities that continually degrade the environment are not sustainable over time-that is, the cumulative impact on the environment of such activities often leads over time to a variety of health, ecological, and other problems. A large portion of the environmental impact in a society is associated with its utilization of energy resources. Ideally, a society seeking sustainable development utilizes only energy resources that cause no environmental impact (e.g., that release no emissions to the environment). Because all energy resources lead to some environmental impact, however, it is reasonable to suggest that some (not all) of the concerns regarding the limitations imposed on sustainable development by environmental emissions and their negative impacts can be overcome in part through increased energy efficiency. Clearly, a strong relationship exists between energy efficiency and environmental impact, because for the same services or products, less resource utilization and pollution normally are associated with increased energy efficiency.

Here, we look at renewable energy resources and compare them with energy conservation. Although not all renewable energy resources are inherently clean, there is such a diversity of choices that a shift to renewables carried out in the context of sustainable development could provide a far cleaner system than would be feasible by tightening controls on conventional energy. Furthermore, being by nature site-specific, they favor power system decentralization and locally applicable solutions more or less independently of the national network. It enables citizens to perceive positive and negative externalities of energy consumption. Consequently, the small scale of the equipment often makes the time required from initial design to operation short, providing greater adaptability in responding to unpredictable growth and/or changes in energy demand.

The exploitation of renewable energy resources and technologies is a key component of sustainable development.[11] There are three significant reasons for it:

- They have much less environmental impact compared with other sources of energy, because there are no energy sources with zero environmental impact. Such a variety of choices is available in practice that a shift to renewables could provide a far cleaner energy system than would be feasible by tightening controls on conventional energy.
- Renewable energy resources cannot be depleted, unlike fossil-fuel and uranium resources. If used wisely in appropriate and efficient applications, they can provide reliable and sustainable supply energy almost indefinitely. By contrast, fossil-fuel and uranium resources are finite and can be diminished by extraction and consumption.
- They favor power system decentralization and locally applicable solutions more or less independently of the national network, thus enhancing the flexibility of the system and the economic power supply to small, isolated settlements. That is why many different renewable energy technologies are potentially available for use in urban areas.

Taking into consideration these important reasons, the relationship between energy conservation and sustainability is finally presented as shown in Fig. 7.

ENERGY CONSERVATION IMPLEMENTATION PLAN

The following basic steps are the key points in implementing an energy conservation strategy plan[11]:

1. *Defining the main goals.* It is a systematic way to identify clear goals, leading to a simple goal-setting process. It is one of the crucial concerns and follows an organized framework to define goals, decide priorities, and identify the resources needed to meet those goals.
2. *Identifying the community goals.* It is a significant step to identify priorities and links among energy, energy conservation, the environment, and other primary local issues. Here, it is also important to identify the institutional and financial instruments.
3. *Performing an environmental scan.* The main objective in this step is to develop a clear picture of the community to identify the critical energy-use areas, the size and shape of the resource-related problems facing the city and its electrical and gas utilities, the organizational mechanisms, and the base data for evaluating the program's progress.
4. *Increasing public awareness.* Governments can increase other customers' awareness and acceptance of energy conservation programs by entering into performance contracts for government activities. They can also publicize the results of these programs and projects. In this regard, international workshops to share experiences on the operation would help overcome the initial barrier of unfamiliarity in countries.
5. *Performing information analysis.* This step carries out a wide range of telephone, fax, email, and Internet interviews with local and international financial institutions, project developers, and bilateral aid agencies to capture new initiatives, lessons learned, and viewpoints on problems and potential solutions.
6. *Building community support.* This step covers the participation and support of local industries and communities, and the understanding the nature of conflicts and barriers between given goals and local actors; improving information flows; activating education and advice surfaces; identifying institutional barriers; and involving a broad spectrum of citizen and government agencies, referring to the participation and support of local industrial and public communities.
7. *Analyzing information.* This step includes defining available options and comparing the possible options with various factors (e.g., program implementation costs, funding availability, utility

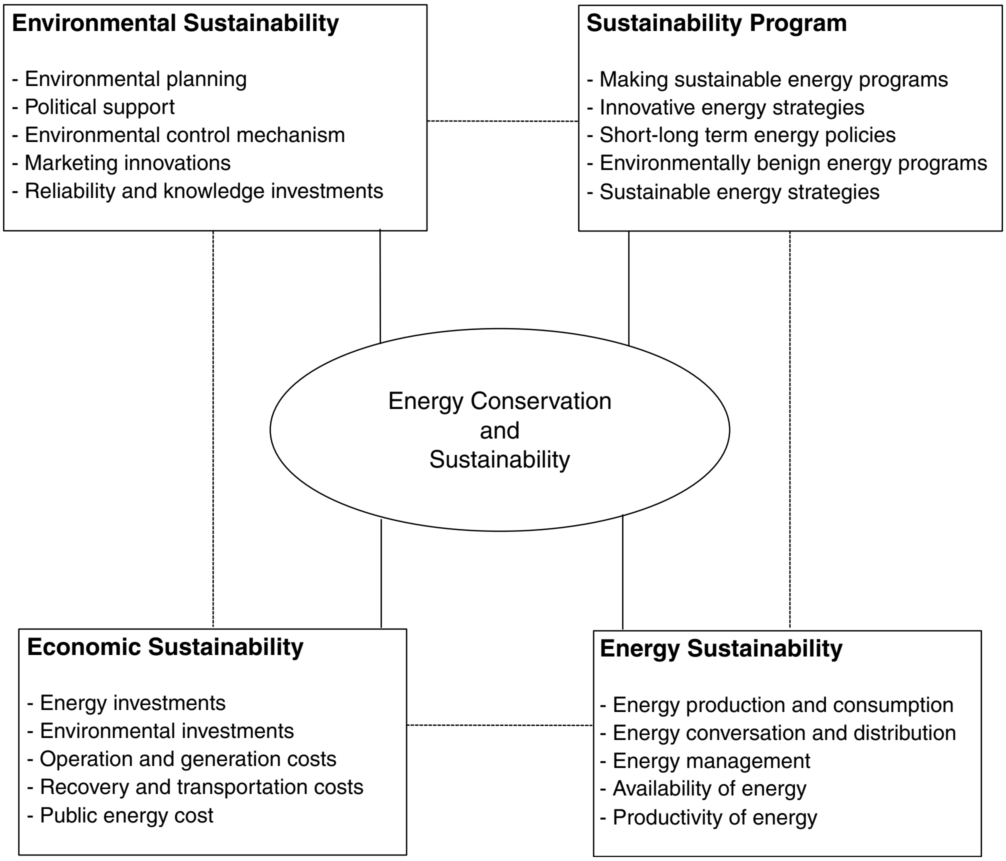


Fig. 7 Linkages between energy conservation and sustainable development.

capital deferral, potential for energy efficiency, compatibility with community goals, and environmental benefits).

8. *Adopting policies and strategies.* Priority projects need to be identified through a number of approaches that are best for the community. The decision-making process should evaluate the cost of the options in terms of savings in energy costs; generation of business and tax revenue; and the number of jobs created, as well as their contribution to energy sustainability and their benefit to other community and environmental goals.
9. *Developing the plan.* When a draft plan has been adopted, it is important for the community to review it and comment on it. The public consultation process may vary, but the aim should be a high level of agreement.
10. *Implementing new action programs.* This step involves deciding which programs to concentrate on, with long-term aims being preferred over short-term aims. The option that has the greatest impact should be focused on, and all details should be defined, no matter how difficult the task seems. Financial resources needed to implement the program should be identified.
11. *Evaluating the success.* The final stage is evaluating and assessing how well the plan performed, which helps identify its strengths and weaknesses and to determine who is benefiting from it.

ENERGY CONSERVATION MEASURES

For energy conservation measures, the information about the measure's applicability, cost range, maintenance issues, and additional points should be presented. Energy conservation involves efficiency improvements, formulation of pricing policies, good "housekeeping practices," and load management strategies, among other measures. A significant reduction in consumer energy costs can occur if conservation measures are adopted appropriately. The payback period for many conservation programs is less than 2 years.

In spite of the potentially significant benefits of such programs to the economy and their proven successes in several countries, conservation programs have not yet been undertaken on a significant scale in many developed and developing countries. Some reasons for this lack of energy conservation programs relate to the following factors:

- Technical (e.g., lack of availability, reliability, and knowledge of efficient technologies)

- Institutional (e.g., lack of appropriate technical input, financial support, and proper program design and monitoring expertise)
- Financial (e.g., lack of explicit financing mechanisms)
- Managerial (e.g., inappropriate program management practices and staff training)
- Pricing policy (e.g., inappropriate pricing of electricity and other energy commodities)
- Information diffusion (e.g., lack of appropriate information)

Reduced energy consumption through conservation programs can benefit not only consumers and utilities, but society as well. In particular, reduced energy consumption generally leads to reduced emissions of greenhouse gases and other pollutants into the environment.

Accelerated gains in energy efficiency in energy production and use, including those in the transportation sector, can help reduce emissions and promote energy security. Although there is a large technical potential for increased energy efficiency, there exist significant social and economic barriers to its achievement. Priority should be given to market forces in effecting efficiency gains. Reliance on market forces alone, however, is unlikely to overcome these barriers. For this reason, innovative and bold approaches are required by governments, in cooperation with industry, to realize the opportunities for energy efficiency improvements, and to accelerate the deployment of new and more efficient technologies.

Here, we look at energy conservation measures, which may be classified in six elements:

- Sectoral measures
- Energy conservation through systematic use of unused energy
- Energy conservation by changing social behavior
- International cooperation to promote energy conservation to counteract global warming
- Enhancing international and government-industry-university cooperation in developing technologies for energy conservation
- Promoting diffusion of information through publicity and education

The emphasis is on sectoral energy conservation. Table 4 presents some examples of such sectoral energy conservation measures. After determining which energy conservation measures are applicable, you should read the description of each of the applicable energy conservation measures. Information about the savings that can be expected from the measure, maintenance issues related to the measure, and other items to consider is provided for each energy conservation measure.

Table 4 Sectoral energy conservation measures

Sector	Measures
Industrial	Strengthening of financial and tax measures to enhance adoption and improvement of energy saving technologies through energy conservation equipment investments
	Re-use of waste energy in factories and/or in surrounding areas
	Enhancing recycling that reduces primary energy inputs such are iron scraps and used papers, and devising measures to facilitate recycling of manufactured products
	Retraining of energy managers and diffusion of new energy saving technologies through them
	Creating database on energy conservation technologies to facilitate diffusion of information
Residential and Commercial	Revising insulation standards provided in the energy conservation law, and introducing financial measures to enhance adoption of better insulation
	Development of better insulation materials and techniques
	Developing' energy conservation' model homes and total energy use systems for homes
	Revising or adopting energy conservation standards for home and office appliances
	Developing more energy saving appliances
	Revising guidelines for managing energy use in buildings, and strengthening advisory services to improve energy management in buildings
Transportation	Because 80% of energy consumption of the sector is by automobiles, further improvement in reducing fuel consumption by automobiles is necessary together with improvement in transportation system to facilitate and reduce traffic flow
	Diffusion of information about energy efficient driving
	Adopting financial measures to enhance the use of energy saving transportation equipment such as wind powered boats

Source: Adapted from Ref. 16.

To evaluate the energy conservation measures, the following parameters should be taken into consideration[13]:

- *Cost estimation.* The first step is to estimate the cost of purchasing and installing the energy conservation measure. Cost estimates should be made for the entire development rather than for a single piece of equipment (e.g., obtain the cost for installing storm windows for an entire development or building, rather than the cost of one storm window). If you are planning to implement the energy conservation measure without the help of an outside contractor, you can obtain cost estimates by calling a vendor or distributor of the product. If, on the other hand, you will be using a contractor to install or implement the energy conservation measure, the contractor should provide estimates that include all labor costs and contract margins.
- *Data survey.* In this step, the questions on fuel consumption and cost should be listed for more than one possible fuel type (e.g., gas, oil, electric, or propane). The appropriate data for each fuel type should be selected and used accordingly for the cost estimation of each fuel.
- *Energy savings.* The amount of energy or fuel used should be estimated.
- *Cost savings.* This step determines the level of savings.
- *Payback period.* The last step in the cost/benefit analysis estimates the simple payback period. The payback period is found by dividing the cost of the measure by the annual cost savings.

LIFE-CYCLE COSTING

The term LCC for a project or product is quite broad and encompasses all those techniques that take into account both initial costs and future costs and benefits (savings) of a system or product over some period of time. The techniques differ, however, in their applications, which depend on various purposes of systems or products. Life-cycle costing is sometimes called a cradle-to-grave analysis. A life-cycle cost analysis calculates the cost of a system or product over its entire life span. Life-cycle costing is a process to determine the sum of all the costs associated with an asset or part thereof, including acquisition, installation, operation, maintenance, refurbishment, and disposal costs. Therefore, it is pivotal to the asset management process.

From the energy conservation point of view, LCC appears to be a potential tool in deciding which system or product is more cost effective and more energy efficient. It can provide information about how to evaluate options concerning design, sites, materials, etc., how to select the best energy conservation feature among various options; how much investment should be made in a single energy conservation feature; and which is the most desirable combination of various energy conservation features.

A choice can be made among various options of the energy conservation measure that produces maximum savings in the form of reduction in the life-cycle costs. A choice can be made between double-glazed and triple-glazed windows, for example. Similarly, a life-cycle cost comparison can be made between a solar heating system and a conventional heating system. The one that maximizes the life-cycle costs of providing a given level of comfort should be chosen. The application of such techniques to energy conservation is related to determining the optimum level of the chosen energy conservation measure. Sometimes, energy conservation measures involve the combination of several features. The best combination can be determined by evaluating the net LCC effects associated with successively increasing amounts of other energy conservation measures. The best combination is found by substituting the choices until each is used to the level at which its additional contribution to energy cost reduction per additional dollar is equal to that for all the other options.

Illustrative Example

Here, we present an illustrative example on LCC to highlight its importance from the energy conservation point of view. This example is a simple LCC analysis of lighting for both incandescent bulbs and compact fluorescent bulbs, comparing their life-cycle costs as detailed in Table 5. We know that incandescents are less expensive (95% to heat and 5% to usable light) and that compact fluorescent bulbs are more expensive but much more energy efficient. So the question is which type of lighting comes out on top in an LCC analysis.

This example clearly shows that LCC analysis helps in energy conservation and that we should make it part of our daily lives.

CONCLUSION

Energy conservation is a key element in sectoral (e.g., residential, industrial, and commercial) energy utilization and is vital for sustainable development. It should be implemented by all possible means, despite the fact that it sometime has its own limitations. This is required not only for us, but for the next generation as well. A secure supply of energy resources is generally considered a necessary but not a sufficient requirement for development within a society. Furthermore, sustainable development demands a sustainable supply of energy resources that, in the long term, is readily and sustainably available at reasonable cost and can be utilized for all required tasks without causing negative societal impact.

Table 5 An example of life-cycle costing (LCC) analysis

Cost of purchasing bulbs	Incandescent	Compact fluorescent
Lifetime of one bulb (hours)	1,000	10,000
Bulb price ($)	0.5	6.0
Number of bulbs for lighting 10,000 h	10	1
Cost for bulbs ($)	10×05=5.0	1×6=6
Energy cost		
Equivalent wattage (W)	75	12
Watt–hours (Wh) required for lighting for 10,000 h	75×10,000=750,000 Wh=750 kWh	12×10,000=120,000 Wh=120 kWh
Cost at 0.05 per kWh	750 kWh×$0.05=$37.5	120 kWh×$0.05=$6
Total cost ($)	5+37.5=42.5	6+6=12

An enhanced understanding of the environmental problems relating to energy conservation presents a high-priority need and an urgent challenge, both to allow the problems to be addressed and to ensure that the solutions are beneficial for the economy and the energy systems.

All policies should be sound and make sense in global terms-that is, become an integral part of the international process of energy system adaptation that will recognize the very strong linkage existing between energy requirements and emissions of pollutants (environmental impact).

In summary this study discusses the current environmental issues and potential solutions to these issues; identifies the main steps for implementing energy conservation programs and the main barriers to such implementations; and provides assessments for energy conservation potentials for countries, as well as various practical and environmental aspects of energy conservation.

REFERENCES

1. Azar, C.; Rodhe, H. Targets for stabilization of atmospheric CO_2. Science **1997**, *276*, 1818–1819.
2. Midilli, A.; Ay, M.; Dincer, I.; Rosen, M.A. On hydrogen and hydrogen energy strategies: I: Current status and needs. Renew. Sust. Energy Rev. **2005**, *9* (3), 255–271.
3. BP. Workbook 2005. *Statistical Review—Full Report of World Energy*, British Petroleum, 2005, Available at: www.bp.com/centres/energy (accessed July 25, 2005).
4. IEE. World Energy Outlook 2002. Head of Publication Services, International Energy, **2002**, Available at: www.worldenergyoutlook.org/weo/pubs/weo2002/WEO21.pdf.
5. Kazman, S. *Global Warming and Energy Policy*; CEI; 67–86, **2003**. Available at: www.cei.org.
6. Dincer, I. Energy and environmental impacts: Present and future perspectives. Energy Sources **1998**, *20* (4/5), 427–453.
7. Dincer, I. *Renewable Energy, Environment and Sustainable Development*. Proceedings of the World Renewable Energy Congress; September 20–25, 1998; Florence, Italy; 2559–2562.
8. Rosen, M.A.; Dincer, I. On exergy and environmental impact. Int. J. Energy Res. **1997**, *21* (7), 643–654.
9. Dincer, I. Renewable energy and sustainable development: A crucial review. Renew. Sust. Energy Rev. **2000**, *4* (2), 157–175.
10. Dincer, I. *Practical and Environmental Aspects of Energy Conservation Technologies*. Proceedings of Workshop on Energy Conservation in Industrial Applications; Dhahran, Saudi Arabia, February 12–14, 2000; 321–332.
11. Dincer, I. On energy conservation policies and implementation practices. Int. J. Energy Res. **2003**, *27* (7), 687–702.
12. Sissine, F. Energy efficiency: Budget, oil conservation, and electricity conservation issues. *CRS Issue Brief for Congress*. IB10020; 2005.
13. Nolden, S.; Morse, D.; Hebert, S. *Energy Conservation for Housing: A Workbook*, Contract-DU100C000018374; Abt Associates, Inc.: Cambridge, MA, 1998.
14. Dincer, I.; Rosen, M.A. A worldwide perspective on energy, environment and sustainable development. Int. J. Energy Res. **1998**, *22* (15), 1305–1321.
15. Painuly, J.P.; Reddy, B.S. Electricity conservation programs: Barriers to their implications. Energy Sources **1996**, *18*, 257–267.
16. Anon. *Energy Conservation Policies and Technologies in Japan: A Survey*, OECD/GD(94)32, Paris, 1994.

Energy Conservation: Industrial Processes

Harvey E. Diamond
Energy Management International, Conroe, Texas, U.S.A.

Abstract
Energy Conservation in Industrial Processes will focus on energy conservation in industrial processes, will distinguish industrial processes and characteristics that differentiate them, will outline the analytical procedures needed to address them, will identify and discuss the main industrial energy intensive processes and some common ways to save energy for each of them, and will address managerial methods of conservation and control for industrial processes.

INTRODUCTION AND SCOPE

Energy conservation is a broad subject with many applications in governmental, institutional, commercial, and industrial facilities, especially because energy costs have risen so high in the last few years and continue to rise even higher. Energy conservation in industrial processes may well be the most important application—not only due to the magnitude of the amount of potential energy and associated costs that can be saved, but also due to the potential positive environmental effects such as the reduction of greenhouse gases associated with many industrial processes and also due to the potential of the continued economic success of all of the industries that provide jobs for many people.

This article will focus on energy conservation in industrial processes—where energy is used to manufacture products by performing work to alter feedstocks into finished products. The feedstocks may be agricultural, forest products, minerals, chemicals, petroleum, metals, plastics, glass, or parts from other industries. The finished products may be food, beverages, paper, wood building products, refined minerals, refined metals, sophisticated chemicals, gasoline, oil, refined petroleum products, metal products, plastic products, glass products, and assembled products of any kind.

This article will distinguish industrial processes and the characteristics that differentiate them in order to provide insight into how to most effectively apply energy conservation within industries. The level of applied technology, the large amount of energy required in many cases to accomplish production, the extreme conditions (e.g., temperature, pressure, etc.) that are frequently required, and the level of controls that are utilized in most cases to maintain process control will be addressed in this article.

This article will outline the analytical procedures needed to address energy conservation within industrial processes and will comment on general analytical techniques that will be helpful in analyzing energy consumption in industrial facilities.

Many of the main energy intensive processes, systems, and equipment used in industries to manufacture products will be identified and discussed in this article and some common ways to save energy will be provided.

This article will cover main energy intensive processes, systems, and equipment in a general format. If more in-depth instruction is needed for explanation of a particular industrial process, system, or type of equipment or regarding the analytical procedures required for a specific process, then the reader should refer to the many other articles included in this Encyclopedia of Energy Engineering and Technology, to references at the Association of Energy Engineers, to the references contained in this article, and if further detail is still needed, then the reader should contact an applicable source of engineering or an equipment vendor who can provide in-depth technical assistance with a specific process, system, or type of equipment.

In addition to the analytical methods of energy conservation, managerial methods of energy conservation will be briefly discussed. The aspects of capital projects versus managerial and procedural projects will be discussed. The justification of managerial efforts in industrial processes will be presented.

INDUSTRIAL PROCESSES—DIFFERENTIATION

Industrial processes require large amounts of energy, sometimes the highest level of technology, and often require very accurate process controls for process specifications, safety, and environmental considerations.

Industrial processes utilize an enormous amount of energy in order to produce the tons of production that are

Keywords: Energy conservation; Industrial processes; Industrial equipment; Process analyses; Process heating; Chemical reactions; Process cooling; Distillation; Drying; Melting and fusion; Mechanical processes; Electrical processes; Energy management.

being produced within industrial facilities. Industrial processes utilize over one-third of the total energy consumed in America.[1] Consider the amount of energy that is required to melt all of the metals being manufactured, to vaporize all of the oil and gasoline being refined, to dry all of the finished products that are made wet, to heat all of the chemicals that react at a certain temperature, to vaporize all of the chemicals that must be distilled for purity, to vaporize all of the steam that is used to heat industrial processes, to mechanically form all of the metal objects that we use, etc.,—this list is too long to be fully included in this article. This is an enormous amount of energy that produces all of the things that humans need and use—food, clothes, homes, appliances, cars, municipal facilities, buildings, roads, etc.

The level of technology required by current industrial processes is the highest in many cases and it is always at a high level in most industrial processes. Most industrial processes are utilizing technology that has been developed in the last 100 years or so, and consequently it has been further improved in the most recent years. Industrial processes most often utilize aspects of chemistry and physics in a precise manner in order to produce the sophisticated products that benefit people in our culture today. Very often, industrial processes require a very high or low temperature or pressure. Often they require a very precise and sophisticated chemistry and commonly they require highly technical designed mechanical processes. The application of electrical equipment and facilities in industrial processes is the highest level of technology for electrical power systems.

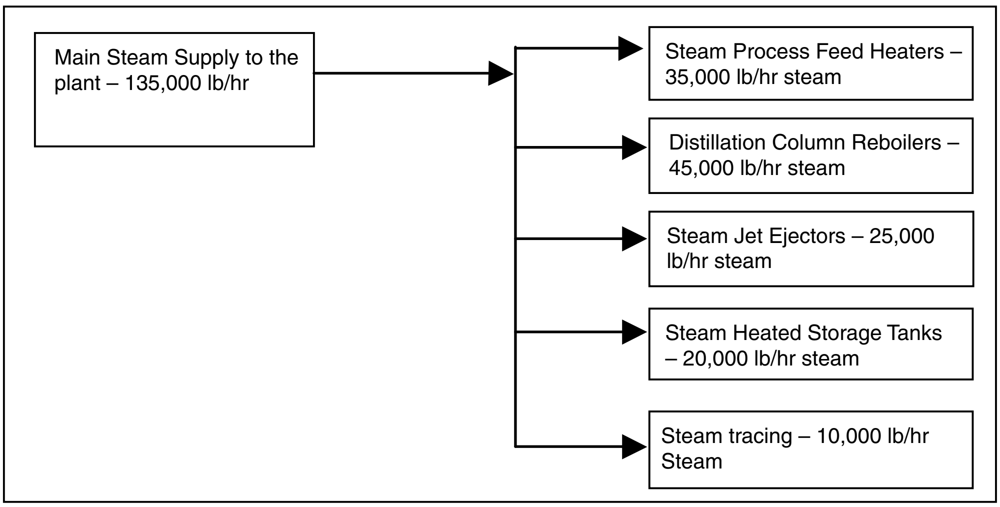
Industrial processes often require the highest level and accuracy of controls in order to produce products that meet product specifications, keep processes operating in a safe manner, and maintain environmental constraints. Due to each of these requirements or due to a combination of these requirements, the process controls for the processes within industrial facilities are often real-time Distributed Control Systems (DCSs), that are of the most sophisticated nature. A typical DCS for industrial processes functions to control process variables instantaneously on a real-time basis, whereby each process variable is being measured constantly during every increment of time and a control signal is being sent to the control element constantly on a real-time basis. The accuracy of a DCS in an industrial facility today is comparable to that of the guidance systems that took the first men to the moon.

Most industrial facilities with DCS controls also utilize a process historian to store the value of most process variables within the facility for a certain increment or period of time. The stored values of these process variables are used for accounting purposes and technical studies to determine optimum operating conditions and maintenance activities.

ENERGY CONSERVATION ANALYSES FOR INDUSTRIAL PROCESSES

In any industrial facility, the first analysis that should be performed for the purpose of energy conservation should be that of determining a balance of the energy consumed for each form of energy. This balance is used to determine how much energy is consumed by each unit, area, or division of the plant or possibly by major items of equipment that are consuming major portions of the energy consumed by the plant. This balance should be determined for each form of energy, whether it is for natural gas, electricity, coal, fuel oil, steam, etc. (see Table 1 below for an example of an energy balance). It

Table 1 Energy balance



might be best that this determination not be called a balance (in that the numbers might not exactly come to a precise balance) but that it sufficiently quantifies the amount of energy consumed by each unit, area, or division of the plant. A better term for this determination might be an "Energy Consumption Allocation". The term balance is more usually applied to chemical and thermodynamic processes where heat and material balances are worked together mathematically to determine a calculated variable and the numbers have to exactly balance in order to arrive at the correct mathematical solution.

Once the amount of energy that is being actually consumed by each part of a plant has been determined, an energy consumption analysis should be performed for each item of energy consuming equipment and each major energy consuming facility in order to determine how much energy should realistically be consumed by each part of the plant. Notice that these calculations are called "realistic" as opposed to just theoretical because the object of these calculations is to determine as closely as possible how much actual energy each item of equipment or part of the plant should be consuming. By comparing these calculations with the actual energy consumption allocations mentioned above, it should be possible to obtain at least an initial indication of where energy is being wasted in a plant (see Table 2 below for an example of an energy consumption analysis for a process feed heater). During the course of obtaining the values of process variables that are required to make these energy consumption analyses, it is possible that indications will be observed of energy wastage due to the presence of an inordinate value of some process variable, such as too high or low of a temperature or pressure. When this type of indication is discovered, it usually also provides insight into what is operating in the wrong way to waste energy. There are numerous instances of energy wastage that can be discovered during these analyses, such as the inordinate manual control of a process, loose operational control of a process variable, or simply not shutting down a piece of equipment when it is not needed.

Table 2 Example of energy consumption analysis: process feed HTR

A process feed heater heats 1,199,520 lb/day of liquid feed material with a specific heat of 0.92 Btu/lb-°F, from 67 to 190°F. A realistic heater efficiency for this type of heater is determined to be 88%. The amount of realistic heat required for this heater is calculated to be: $Q = 1{,}199{,}520$ lb/day × 0.92 Btu/lb-°F × (190°F − 67 °F) ÷ 0.88 = 154,247,367.3 Btu/day of realistic heat consumption.
It is observed that this feed heater is consuming 186,143,720 Btu/day
This feed heater is being operated in a wasteful way and is wasting over 20% of its heat

The analyses discussed in the above paragraph encompass all technical engineering science subjects, such as chemistry, thermal heat transfer, thermodynamics, fluid mechanics, mechanical mechanisms, and electrical engineering.

The next set of energy conservation analyses that should be performed are used to calculate the efficiencies of each item of equipment or facility to which efficiency calculations would be applicable, such as boilers, fired heaters, furnaces, dryers, calciners, and all other thermodynamic processes (efficiency calculations for boilers and other combustion equipment is available in the *Energy Management Handbook* by Wayne C. Turner and Steve Doty[2] and in the *Guide to Energy Management* by Barney L. Capehart, Wayne C. Turner, and William J. Kennedy)[3] and for electrical and mechanical equipment such as motors, pumps, compressors, vacuum pumps, and mechanical machinery. Once the actual efficiencies of any of the above have been determined, these numbers can be compared to the realistic efficiency for the type of equipment or facility that is prevalent throughout industry. These calculations and comparisons will also reveal wastage of energy and will frequently identify the causes of energy wastage and the possible issues to be corrected.

The next level of energy conservation analysis that may be performed is process analysis that can be conducted on a particular chemical, thermodynamic, thermal, fluid flow, mechanical, or electrical process. These analyses are usually performed by experienced engineers to examine the process itself and the process variables to determine if the process is being operated in the most effective and efficient manner. Here again, an indication will be provided as to whether or not energy is being wasted in the actual operation of the process. Chemical, thermodynamic, thermal, fluid flow, and other processes, as well as combinations of any of these processes can often require process simulation software such as PROMAX by Bryan Research & Engineering, Inc.,[4] in order to properly analyze these processes. The analysis of distillation columns, evaporators, and dryers can fall into this category. A good example would be the process analysis of a distillation column to determine if an effective and efficient level of reflux to the column and reboiler duty is being used.

Another analysis that has been very useful in the past few years in identifying energy conservation projects is Pinch analysis. This analysis is performed on thermodynamic and thermal processes in order to identify sources of energy within existing processes that can be used to supply heat for these processes instead of having to add additional heat to the entire process. The net effect is to reduce the amount of energy required for the overall process. The performance of a Pinch analysis on a particular process or facility will usually identify capital projects where revisions to the facility can be made to decrease the total amount of energy required. These are very often waste heat recovery projects. See "Use Pinch Analysis to Knock Down Capital Costs and Emissions" by Bodo Linnhoff, Chemical Engineering, August 1994[5] and "Pinch Technology: Basics for the Beginners".[6]

MAIN INDUSTRIAL ENERGY PROCESSES, SYSTEMS, AND EQUIPMENT

This section provides an overview and a list of the more common energy intensive industrial processes that are used to manufacture products in industrial facilities. Most energy intensive industrial processes can be classified into about eight general process categories—process heating, melting, chemical reactions, distillation-fractionation, drying, cooling, mechanical processes, and electrical processes. These processes are intended to be the main general energy intensive processes that are most commonly used and to which variations are made by different industries in order to make a specific product. In this regard, this is an overview—these processes are often not the specific process but a general category to which variations can be made to achieve the specific process.

In the following paragraphs, each process will be discussed by addressing its description, what systems it utilizes, what products are generally made, how it uses energy, and frequent ways that energy can be saved.

Common energy consuming systems and equipment that work to manufacture products in industrial facilities are also listed below and discussed in the same manner as the main industrial processes, as they are also common to industrial facilities and are related to these processes.

Process Heating

Description. The addition of heat to a target in order to raise its temperature. Temperatures can range from the hundreds to the thousands in industrial process heating.
Energy form. Heat must be generated and transferred to the intended object or medium.
Energy unit. Btu, calorie, joule, therm or watt-hour.
Examples. The application of heat in order to heat feed materials, to heat chemical processes, to heat metals for forming, to heat materials for drying, to heat materials in a kiln or calciner, to heat minerals and metals for melting.
Applied systems. Combustion systems, steam systems, thermal systems, hot oil systems, heating medium systems such as Dowtherm[7] or Paracymene,[8] and electrical resistance or induction heating systems.
Common equipment. Boilers, furnaces, fired heaters, kilns, calciners, heat exchangers, waste heat recovery exchangers, preheaters, electrical resistance heaters, and electrical induction heaters.
Common energy conservation issues. Keeping the heat targeted at the objective—proper insulation, seals on enclosures, eliminating leakage, and eliminating unwanted air infiltration. Control issues—maintaining sufficient control of the heating process, temperatures, and other process variables to avoid waste of heat. Management issues—shutting down and starting up heating processes at the proper times in order to avoid waste of heat and management of important process variables to reduce the amount of heat required to accomplish the proper process. Application of Pinch Technology—identify process areas where heat can be recovered, transferred, and utilized to reduce the overall process heat requirement. Waste heat recovery.

Melting and Fusion

Description. The addition of heat or electrical arc energy at a high temperature in order to melt metals, minerals, or glass. The melting process involves more than just process heating, it involves fluid motion, fluid density equilibrium, chemical equilibrium, cohesion, and sometimes electro-magnetic inductance. Reference: "The study showed that the fluid equations and the electromagnetic equations cannot be decoupled. This suggests that arc fluctuations are due to a combination of the interactions of the fluid and the electromagnetics, as well as the rapid change of the boundary conditions."[9]
Energy forms. Heat at high temperatures or electrical arc energy in the form of high voltage and high current flowing in an arc.
Energy units. Heat, Btu, calorie, joule, or therm.
Electrical arc. KWhrs.
Examples. Melting of ores in order to refine metals such as iron, aluminum, zinc, lead, copper, silver, etc. Melting of minerals in order to refine minerals such as silica compounds, glass, calcium compounds, potassium alum, etc.
Applied systems. Combustion systems, chemical reactions, and electrical systems.
Common equipment. Blast furnaces, arc furnaces, electrical resistance heaters, and electrical induction heaters.
Common energy conservation issues. Pre-condition of feed material—moisture content, temperature, etc. Feed method—efficiency of melting process effected by the feed method, feed combinations, and feed timing. Control of electromagnetics during melting and use of magnetic fields during separation. Overheating can waste energy without yielding positive process results. Heat losses are due to poor insulation, the failure of seals, or lack of shielding or enclosure.

Chemical Reactions

Description. Chemicals react to form a desired chemical, to remove an undesired chemical, or to break out a desired chemical. The chemical reaction can involve heat, electrolysis, catalysts, and fluid flow energy.
Energy forms. Heat, electrolysis, and fluid flow.
Energy units. Heat, Btu, calorie, joule, or therm.
Electrolysis. KWhrs.
Fluid flow. Ft-Lbs or Kg-M.
Examples. Reaction of chemical feed stocks into complex chemicals, petrochemical monomers into polymers, the oxidation of chemicals for removal, dissolving of chemicals to remove them, reaction of chemicals with other chemicals to remove them, the reaction of lignin with reactants in order to remove it from cellulose, the electro-plating of metals out of solution to refine them.
Applied systems. Feed systems, catalysts systems, heating systems, cooling systems, vacuum systems, run-down systems, separation systems, filtering systems, and electro-plating systems.
Common equipment. Reactors, digesters, kilns, calciners, smelters, roasters, feed heaters, chillers, pressure vessels, tanks, agitators, mixers, filters, electrolytic cells.
Common energy conservation issues. Close control of heating and cooling for chemical reactions. Close control of all reaction process variables—balance of all constituents, amount of catalyst, proper timing on addition of all components. Management of feedstocks, catalysts, and run-down systems for proper timing and correct balance for highest efficiency. Pinch analysis of feed heating, run-down products, cooling system, etc. Conservation of heating and cooling—proper insulation, sealing, and air infiltration. Waste heat recovery.

Distillation-Fractionation

Description. A thermo-dynamic and fluid flow equilibrium process where components of a mixture can be separated from the mix due to the fact that each component possesses a different flash point.
Energy form. Heat and fluid flow.
Energy units. Heat, Btu, calorie, joule, or therm.
Fluid flow. Ft-Lbs or KG-M.
Examples. Distillation-fractionation of hydrocarbons in oil and gas refineries and chemical plants. Distillation of heavy hydrocarbons in gas processing plants where natural gas is processed to remove water and heavy hydrocarbons.
Applied systems. Feed heating systems, over-head condensing systems, reflux systems, reboil systems, vacuum systems.
Common equipment. Distillation columns or towers, over-head condenser heat exchangers and accumulators—vessels, reflux pumps, reboiler heat exchangers, feed pumps, feed—effluent heat exchangers, vacuum steam jet ejectors.
Common energy conservation issues. Feed temperatures, reflux ratios, reboiler duty. Close control on pressures, temperatures, feed rates, reflux rates, and reboil duty. Management of overall operation timing—running only when producing properly. Concurrent use of vacuum systems—only when needed. Pinch analysis for feed and effluent streams and any process cooling systems. Proper insulation and elimination of lost heat for fired heater reboilers.

Drying

Description. The use of heat and fluid flow to remove water or other chemical components in order to form a more solid product.
Energy forms. Heat and fluid flow.
Energy units. Heat, Btu, calorie, joule, or therm.
Fluid flow. Ft-Lbs or KG-M.
Examples. Spray dryers that dry foods, sugar, fertilizers, minerals, solid components, and chemical products. Rotary dryers that dry various loose materials. Line dryers that dry boards, tiles, paper products, fiberglass products, etc. Other dryers that dry all kinds of products by flowing heated air over finished products in an enclosure.
Applied systems. Combustion systems, steam systems, thermal heating systems, cyclone systems, air filter systems, incinerator systems, Regenerative Thermal Oxidizer (RTO) systems.
Common equipment. Spray dyers, spray nozzles, natural gas heaters, steam heaters, electrical heaters, blowers, fans, conveyors, belts, ducts, dampers.
Common energy conservation issues. Efficient drying process for the components being eliminated. Proper amount of air flowing through dryer for drying. Proper insulation, seals, and elimination of lost heat due to infiltration. Waste heat recovery.

Process Cooling

Description. The removal of heat by a cooling medium such as cooling water, chilled water, ambient air, or direct refrigerant expansion.
Energy form. Heat.
Energy unit. Btu, calorie, joule, or therm.
Examples. Cooling water or chilled water circulated through a cooling heat exchanger, an air cooled heat exchanger, or a direct expansion evaporator that cools

air for process use.
Applied systems. Cooling water systems, chilled water systems, refrigerant systems, thermal systems.
Common equipment. Cooling towers, pumps, chillers, refrigeration compressors, condensers, evaporators, heat exchangers.
Common energy conservation issues. Use evaporative cooling as much as possible. Keep chillers properly loaded. Restrict chilled water flow rates to where 10°F temperature difference is maintained for chilled water. Limit cooling water pumps to the proper level of flow and operation. Apply Pinch analysis to achieve most efficient overall cooling configuration. Proper insulation, seals, and elimination of air infiltration.

Mechanical Processes

Description. Physical activities that involve force and motion that produce finished products. Physical activities can be discrete or can be by virtue of fluid motion.
Energy form. Physical work.
Energy unit. Ft-Lbs or KG-M.
Examples. Machining of metals, plastics, wood, etc.; forming or rolling or pressing of metals, minerals, plastics, etc.; assembly of parts into products; pumping of slurries thru screens or filters for separation; cyclone separation of solids from fluids; pneumatic conveyance systems that remove and convey materials or products and separate out solids with screens or filters.
Applied systems. Machinery, electrical motors, hydraulic systems, compressed air systems, forced draft or induced draft conveyance systems, steam systems, fluid flow systems.
Common equipment. Motors, engines, turbines, belts, chains, mechanical shafts, bearings, conveyors, pumps, compressors, blowers, fans, dampers, agitators, mixers, presses, moulds, rolls, pistons, centrifuges, cyclones, screens, filters, filter presses, etc.
Common energy conservation issues. Equipment efficiencies, lubrication, belt slippage, hydraulic system efficiency, compressed air system efficiency. Control of process variables. Application of variable speed drives and variable frequency drives. Management of system and equipment run times.

Electrical Processes

Description. The application of voltage, current, and electromagnetic fields in order to produce products.
Energy form. Voltage-current over time; electromagnetic fields under motion over time.
Energy units. KWh.
Examples. Arc welding, arc melting, electrolytic deposition, electrolytic fission, induction heating.
Applied systems. Power generator systems, power transmission systems, amplifier systems, rectifier systems, inverter systems, battery systems, magnetic systems, electrolytic systems, electronic systems.
Common equipment. Generators, transformers, relays, switches, breakers, fuses, plates, electrolytic cells, motors, capacitors, coils, rectifiers, inverters, batteries.
Common energy conservation issues. Proper voltage and current levels, time intervals for processes, electromagnetic interference, hysteresis, power factor, phase balance, proper insulation, grounding. Infrared scanning of all switchgear and inter-connections.

Combustion Systems

Combustion systems are found in almost all industries in boilers, furnaces, fired heaters, kilns, calciners, roasters, etc. Combustion efficiency is most usually a prime source of energy savings.

Boilers and Steam Systems

Boilers and steam systems may well be the most widely applied system for supplying process heat to industrial processes. "Over 45% of all the fuel burned by U.S. manufacturers is consumed to raise steam."[10] Boiler efficiencies, boiler balances (when more than one boiler is used), and steam system issues are usually a prime source of energy savings in industrial facilities.

Flare Systems and Incinerator Systems

Flare and incinerator systems are used in many industrial facilities to dispose of organic chemicals and to oxidize volatile organic compounds. Proper control of flares and incinerators is an issue that should always be reviewed for energy savings.

Vacuum Systems

Vacuum systems are used to evaporate water or other solvents from products and for pulling water from products in a mechanical fashion. Vacuum systems are also used to evacuate hydrocarbon components in the petroleum refining process. Vacuum systems are frequently used in the chemical industry to evacuate chemical solvents or other components from a chemical process. Steam jet ejectors and liquid ring vacuum pumps are commonly used to pull vacuums within these systems. The efficiencies of the ejectors and the liquid ring vacuum pumps can be a source of energy savings as well as the management of vacuum system application to production processes. Pneumatic conveyance systems that utilize a fan or blower to create a low-level vacuum are sometimes used to withdraw materials or products from a process and separate the matter within a screen or filter. For large conveyance systems, the efficiencies of the equipment and the management of their operation can be a source of energy savings.

Furnaces, Fired Heaters, Kilns, Calciners

The above comments on combustion systems are applicable to these equipment items and additional energy savings issues can be found relative to them.

Centrifugal Pumps

Centrifugal pumps are used widely in industries. The flow rate being pumped is a primary determining factor for the amount of power being consumed and it is sometimes higher than required. Good control of the pumping rate is an important factor in saving energy in centrifugal pumps. The application of variable frequency drives to the motor drivers can be a good energy saving solution for this issue.

Fans and Blowers

The flow rate for fans and blowers is analogous to the pumping rate above for centrifugal pumps. Good control of the flow rate and the possible application of Variable Frequency Drive (VFD) apply here as well for fans and blowers.

Centrifugal Compressors

Compressors are used widely in industry. The above discussions of flow rates, control of flow rates, and application of VFDs apply here as well. Centrifugal compressors frequently will have a recycle flow that is controlled in order to prevent the compressor from surging. Close control of this recycle flow at its minimum level is very important for compressor efficiency.

Liquid Ring Vacuum Pumps

As mentioned above in several places, liquid ring vacuum pumps are used widely in industry. The amount of sealing liquid that is recycled to the pump and the temperature of the sealing liquid are important determinates of the efficiency of the Liquid Ring Vacuum Pump (LRVP).

The above overview and list of industrial processes, systems, and equipment has been general in nature due to the limitations of this article. Greater and more specific familiarity with each of these industrial energy intensive processes, systems, and equipment will yield greater applicable and effective insight into ways to save energy related to each of these items.

CAPITAL PROJECTS VERSUS IMPROVED PROCEDURES

Energy conservation effort applied in industrial facilities can identify capital projects whereby the facilities can be changed in order to achieve greater overall energy efficiency or the efforts can identify changes to in day-to-day operating and maintenance procedures that can reduce waste of energy and also improve the overall efficiency of the facility. Frequently, energy-saving procedural changes to day-to-day operations and maintenance activities within an industrial plant can be identified by taking and recording operating data once the processes, systems, and equipment have been studied and analyzed for energy consumption. Procedural changes to operations and maintenance within an industrial plant can often amount to low costs or possibly no costs to the facility. This aspect of energy conservation is often overlooked by highly technical personnel that have worked hard to design industrial facilities because they have technically designed the facility very well for energy consumption considerations and the more mundane activities related to day-to-day operation and maintenance tend to not register in their highly technical perspective. None-the-less, a considerable amount of energy can usually be saved within most industrial processes, systems, and equipment due to changes in the way they are operated and maintained. A general tendency within industrial plants is that operations will often operate the processes and systems at a point that provides a comfortable separation between an operating variable and its limitation in order to understandably ensure no upsets occur within the process or system. However, with the cost of energy being what it is today, it is frequently found that a significant amount of energy can be saved by operating processes and systems more tightly and efficiently, even though it may require more attention, increased control, and the monitoring of process variables.

EFFECTIVE ENERGY MANAGEMENT SYSTEMS

Another aspect of energy conservation that can be very productive in saving energy within industrial processes is that of an effective energy management system. An effective energy management system is comprised of operational and maintenance managers functioning in conjunction with an accurate and concurrent data collection system in order to eliminate waste and improve overall efficiency of industrial processes. It is not possible to manage any activity unless the activity is being properly monitored and measured with key performance metrics (KPMs). The data collection system part of an effective energy management system within any industrial facility provides the accurate and concurrent measurement data (KPMs) that is required in order to identify actions that are needed to eliminate waste of energy and improve overall efficiency of the facility. An effective energy management system is first built upon acquiring total knowledge of the facility down into every level of operation and maintenance of the facility. Such a level of thoroughness and complete analysis of energy consumption within a facility is sometimes referred to as *Total Quality Energy* management.[11] Once an effective energy management system has been established and is effectively controlling energy consumption of an industrial facility, it should be maintained, in effect, so that it will continue to monitor KPMs to maintain energy conservation for the facility. An effective energy management system within an industrial manufacturing facility can eliminate as much as two to three percent of the energy costs by eliminating waste of energy on a day-to-day operational and maintenance basis. In most industrial facilities, this level of cost reduction is significant and will justify an effective energy management system.

THE CURRENT NEED FOR GREATER ENERGY CONSERVATION IN INDUSTRY

With the present cost of all forms of energy today, it would certainly seem logical that all of industry would be seeking greater energy conservation efforts within their facilities. Unfortunately, many corporate industrial managers are not aware of the true potential of conserving energy within their processes and facilities. Greater awareness of the ability to conserve energy on the basis of increased efficiencies of processes, systems, and equipment is needed; and also due to the application of an effective energy management system. For the good of society and environment, corporate industrial managers should be more open to the possibility of the improvement of industry that will work to sustain their business and improve the world that we live in. This is in opposition to corporate political thinking, which does not want to consider making changes and wants no one to interfere with their present activities. Human beings should be willing to examine themselves and make changes that will make things better. The same outlook should be applied to businesses and industry in order to make things better. Greater management support is needed in industry today to accomplish greater and very much needed increased energy conservation.

CURRENT APPLICATION OF INCREASED ENERGY MANGEMENT

With the recent technological advancements that have been made in digital computer and communications systems, data collection systems can be implemented in industrial facilities in a much more cost effective manner. Wireless communication systems for metering and data collection systems have advanced dramatically in the last few years and network-based computer communication has enabled whole new systems for measurement and control. With all of these new fields of configuration for data collection systems, with the increased technology, and with the lower costs to accomplish data collection systems, it is now possible to apply energy management systems to industry today with much greater applicability. Hopefully this will be recognized and result in greater applications of effective energy management systems.

From recent observations, it appears that most of industry today is a candidate for improved and more effective energy management systems. In conjunction with the increased technology and lower cost potentials, it seems that there is a definite match between supply and need for the application of increased energy management systems.

SUMMARY

Industrial processes have commonality in processes, systems, and equipment. There are logical and systematic analyses that can be performed in industrial processes that can identify ways to save energy. Effective energy management systems are needed in industry today and there are great possibilities to save energy in industrial processes. Energy can be conserved in industrial processes by analyses that will improve efficiencies, by implementation of procedures that eliminate waste, and by application of an effective energy management system.

REFERENCES

1. U.S. Department of Energy–Energy Efficiency and Renewable Energy, http://www.eere.energy.gov/EE/industry.html (accessed on 2006).
2. Turner, W.C.; Doty, S. *Energy Management Handbook*, 6th Ed.; Fairmont Press: Lilburn, GA, 2005.
3. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 5th Ed.; Fairmont Press: Lilburn, GA, 2004.
4. Bryan Research & Engineering, Inc. PROMAX; BRE, Bryan, TX; http://www.bre.com (accessed on 2006).
5. Linnhoff, B. Use pinch analysis to knock down capital costs and emissions. Chemical Engineering **August 1994** http://www.che.com.

6. Solar Places Technology. Pinch Technology: Basics for the Beginners; http://www.solarplaces.org/pinchtech.pdf (accessed on 2006).
7. Dowtherm, Dow Chemical http://www.dow.com/heattrans/index.html (accessed on 2006).
8. Paracymene, Orcas International, Flanders, NJ 07836 http://www.orcas-intl.com (accessed on 2006).
9. King, P.E. Magnetohydrodynamics in Electric Arc Furnace Steelmaking. Report of Investigations 9320; United States Department of the Interior, Bureau of Mines; http://www.doi.gov/pfm/ar4bom.html (accessed on 2006).
10. U.S. Department of Energy–Energy Efficiency and Renewable Energy, http://www.eere.energy.gov/EE/industry.html the common ways (accessed on 2006).
11. Energy Management International, Inc. Total Quality Energy; http://www.wesaveenergy.com (accessed on 2006).

Energy Conservation: Lean Manufacturing

Bohdan W. Oppenheim
U.S. Department of Energy Industrial Assessment Center, Loyola Marymount University, Los Angeles, U.S.A.

Abstract
Productivity has a major impact on energy use and conservation in manufacturing plants—an impact often more significant than optimization of the equipment energy efficiency. This article describes Lean Manufacturing, which represents the current state-of-art in plant productivity. A significant opportunity for energy savings by transforming production into single-piece Lean Flow is demonstrated. The impact of major individual productivity elements on energy is discussed. Simple metrics and models are presented as tools for relating productivity to energy. Simple models are preferred because productivity is strongly influenced by intangible human factors such as work organization and management, learning and training, communications, culture, and motivation, which are difficult to quantify in factories.

INTRODUCTION

At the time of this writing (2005), the world is experiencing strong contradictory global trends of diminishing conventional energy resources and rapidly increasing global demands for these resources, resulting in substantial upwards pressures in energy prices. Because the energy used by industry represents a significant fraction of the overall national energy use, equal to 33% in the United States in the year 2005, a major national effort is underway to conserve industrial energy.[1] The rising energy prices place escalating demands on industrial plants to reduce energy consumption without reducing production or sales, but by increasing energy density.

Optimization of industrial hardware and its uses, including motors and drives, lights, heating, ventilation and cooling equipment, fuel-burning equipment, and buildings, are well understood, have been practiced for years,[2] and are important in practice. However, they offer only limited energy conservation opportunities, rarely exceeding a few percent of the preoptimization levels. In contrast, the impact of productivity on energy use and energy density offers dramatically higher savings opportunities in energy and in other costs. In the extreme case, when transforming a factory from the traditional "process village" batch-and-queue system to the state-of-the-art, so-called Lean system, the savings in energy can reach 50% or more.

The best organization of production known at this time is called Lean, developed at Toyota in Japan.[3] It is the flow of value-added work through all processes required to convert raw materials to the finished products with minimum waste. Major elements of Lean organization include: steady single-piece flow with minimum inventories and no idle states or backflow; flexible production with flexible equipment and operators and flexible floor layouts ready to execute the order of any size profitably and just-in-time; reliable and robust supplies of raw materials; minimized downtime due to excellent preventive maintenance and quick setups; first-pass quality; clean, uncluttered, and well-organized work space; optimized work procedures; and, most importantly, an excellent workforce–well trained, motivated, team-based and unified for the common goals of having market success, communicating efficiently, and being well-managed. The Lean organization of production is now well understood among productivity professionals, but it is not yet popular among the lower tier suppliers in the United States. Its implementation would save energy and benefit the suppliers in becoming more competitive.

The engineering knowledge of energy conservation by equipment improvements is well understood and can be quantified with engineering accuracy for practically any type of industrial equipment.[2] In contrast, industrial productivity is strongly influenced by intangible and complex human factors such as management, work organization, learning and training, communications, culture, and motivation. These work aspects are difficult to quantify in factory environments. For this reason, the accuracy of productivity gains and the related energy savings are typically much less accurate than the energy savings computed from equipment optimization. Simple quantitative models with a conservative bias are therefore recommended as tools for energy management in plants. This article includes some examples. They are presented in the form of energy savings or energy cost savings that would result from implementing a given productivity

Keywords: Productivity impact on energy; Productivity; Energy; Lean; Just-in-time; Energy savings; Energy conservation; Energy density; Industrial energy.

improvement, or eliminating a given productivity waste, or as simple metrics measuring energy density.

It is remarkable that in most cases, these types of energy savings occur as a natural byproduct of productivity improvements, without the need for a direct effort centered on energy. Thus, the management should focus on productivity improvements. In a traditional non-Lean plant intending to transform to Lean production, the first step should be to acquire the knowledge of the Lean system. It is easily available from industrial courses and workshops, books,[3,4] and video training materials.[6] The next step should be the actual transformation of production to Lean. Most of the related energy savings will then occur automatically. Implementation of individual productivity elements such as machine setup time reduction will yield some energy savings, but the result will not be as comprehensive as those yielded by the comprehensive implementation of Lean production.

TRADITIONAL VS LEAN PRODUCTION

The traditional organization of production still used frequently in most factories tends to suffer from the following characteristics:

- Supplier selection is based on minimum cost, resulting in a poor level of mutual trust and partnership, the need for receiving inspection, and often large inventories of raw materials (RM).
- Work-in-progress (WIP) is moving in large batches from process village to process village and staged in idle status in queues in front of each machine, while the machine moves one piece at a time. This work organization is given the nickname "batch-and-queue" (BAQ).[3]
- Finished goods (FG) are scheduled to complex forecasts rather than customer orders, resulting in large inventories.
- The floor is divided into "process villages" populated with large, complex, and fast similar machines selected for minimum unit cost.
- Minimum or no information is displayed at workstations, and the workers produce quotas.
- Work leveling is lacking, which results in a random mix of bottlenecks and idle processes.
- Unscheduled downtime of equipment occurs frequently.
- Quality problems with defects, rework, returns, and customer complaints are frequent.
- Quality assurance in the form of 100% final inspections attempts to compensate for poor production quality.
- The floor space is cluttered, which makes moving around and finding items difficult.
- The workforce has minimum or no training and single skills.
- The management tends to be authoritarian.
- A language barrier exists between the workers and management.
- There is a culture of high-stress troubleshooting rather than creative trouble prevention.

In such plants, the waste of materials, labor, time, space, and energy can be as much as 50%–90%.[3]

The Lean production method developed primarily at Toyota in Japan under the name Just-In-Time (JIT), and generalized in the seminal work[3] is the opposite of the traditional production in almost all respects, as follows:

- Raw materials are bought from reliable supplier–partners and delivered JIT in the amount needed, at the price agreed, and with the consistently perfect quality that obviates incoming inspection.
- Single-piece flow (SPF) of WIP is steadily moving at a common takt time (Takt time is the common rhythm time of the pieces moving from workstation to workstation on the production line. It is the amount of time spent on EACH operation. It precisely synchronizes the rate of all production operations to the rate of sales JIT.), from the first to the last process.
- The FG are produced to actual customer orders JIT resulting in minimum inventories.
- The floor is divided into flexible production lines with small simple machines on casters that can be pushed into position and setup in minutes.
- The labor is multiskilled, well motivated and well trained in optimized procedures.
- Quality and production status are displayed on large visible boards at each workstation, making the entire production transparent for all to see.
- Preventive maintenance assures no unscheduled downtime of equipment.
- All process operators are trained in in-line quality checks and variability reduction.
- No final inspection is needed, except for occasional sampled checks of FG.
- Defects, rework, returns, and customer complaints are practically eliminated.
- The floor space is clean and uncluttered.
- The workforce is trained in company culture and commonality of the plant mission, customer needs, workmanship, and quality.
- The culture promotes teamwork, multiple job skills, supportive mentoring management, and company loyalty.
- The management promotes trouble prevention and "stopping the line" at the first sign of imperfection so that no bad pieces flow downstream.

According to Womack et al. the transformation from traditional to Lean production can reduce overall cost,

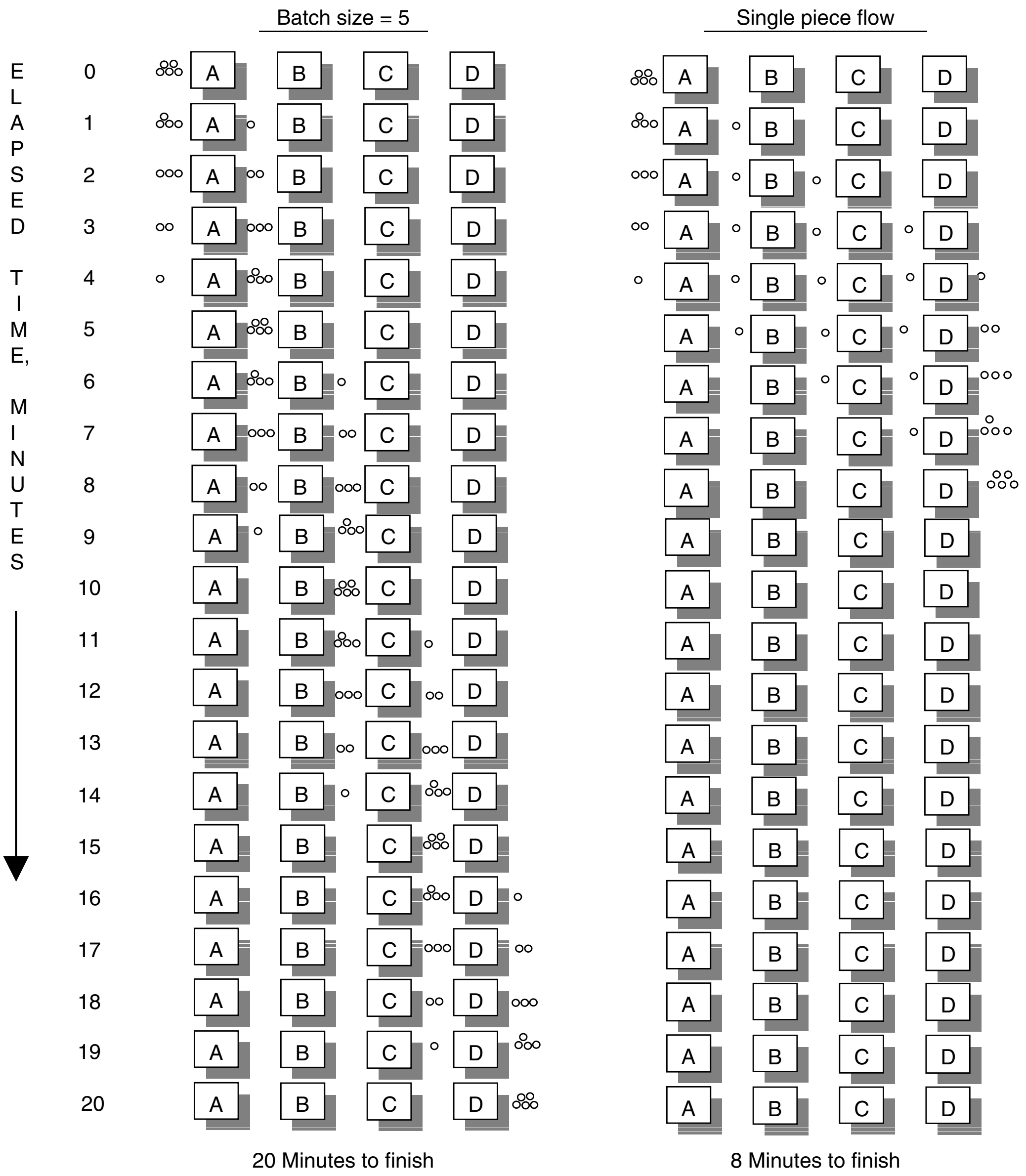


Fig. 1 BAQ with batch size of five vs SPF.

inventory, defects, lead times by 90%, and space by 50%, and vastly increase plant competitiveness, customer satisfaction, and workforce morale. The resultant energy savings can be equally dramatic. Ref. [4] contains interviews with industry leaders who have succeeded in this transformation.

IMPACT ON ENERGY

The impact of productivity on plant energy falls into the following two broad categories:

1. Productivity improvements that save infrastructure energy. These improvements reduce the energy consumed by all plant support systems, which tend to be energized regardless of the actual production activities, such as lights, space cooling and heating devices, cooling towers, combustion equipment (boilers, molten metal furnaces), air compressors, forklift battery chargers, conveyors, etc. To the first approximation, the infrastructure energy is reduced in proportion to the production time reductions, which can be huge in the Lean system. In order to perform more detailed estimates of the infrastructure energy savings, the management would have to conduct detailed energy accounting and understand how much energy is used by each support system under different production conditions. This knowledge is rarely available;

therefore the former simplistic approach, combined with conservative estimates, offer useful tools.

2. Process energy savings. In this category, the energy savings of process equipment are obtained by improving the process productivity. Examples include the reduction of unscheduled machine downtime or setup time and the elimination of process variability, defects, rework, scrap, excessive labor time, etc.

Single Piece Flow (SPF)

Changing the traditional BAQ production to Lean production is by far the most effective productivity transformation a plant can undertake, creating dramatic savings in the overall throughput time, cost, quality, and energy. The example shown in Fig. 1 compares just one aspect of the transformation—a reduction of batch size from five to one, i.e., the SPF. In both cases, four processes of equal one-minute takt time are assumed. The benefits of the SPF alone are dramatic, as follows:

1. In BAQ, the batch is completed in 20 min and in SPF in only 8 min, a 60% reduction.
2. In BAQ, only one machine at a time produces value, while three others are idle. If the idle machines remain energized, as is the case, e.g., with injection molding, three of the four machines (75%) would be wasting energy, and doing it for 16 min each, adding up to 64 min of machine energy wasted. In the SPF system, no machine energy is wasted as no machine would be idle, except for the lead and tail of each process of 4 min, adding up to 16 min of machine energy wasted, a savings of 75% from BAQ.
3. Reducing the batch throughput time by 60% reduces the infrastructure energy by the same amount, assuming the production is completed faster and the plant is de-energized. Alternatively, the freed 60% time and energy could be used for additional production and profits.
4. An important additional benefit is that in SPF, a defect can be detected on the first specimen–as soon as it reaches the next process, while in the BAQ, the entire batch may be wasted before the defect is discovered and a corrective action undertaken, with the energy used for making the batch wasted.

This simple example clearly illustrates the dramatic impact of SPF on both overall productivity and energy consumption. Typically, as the factories transform to the Lean system, their sales, production, and profits increase simultaneously and the energy used decreases. A convenient metric to track the overall benefit is the gross energy density, ED_1 or ED_2:

$$ED_1 = \frac{EC_T}{P} \tag{1a}$$

$$ED_2 = \frac{EC_T}{AC} \tag{1b}$$

where EC_T is the overall annual cost of energy in the plant, P is the number of products produced per year and AC is the total annual costs (sales minus net profit). ED_1 should be used if similar products are made most of the time, and ED_2 should be used if the plant has a wide menu of dissimilar products. The ED ratios will decrease as progress is made from BAQ to SPF. If the volume of production remains constant during the transformation, energy savings and energy cost savings alone may be more convenient metrics to track plant energy efficiency.

Inventory Reduction

All inventories, whether in RM, WIP, or FG, beyond the immediate safety buffers, are detrimental. Inventory means that company capital is "frozen" on the floor; cutting into the cash flow; wasting labor for inventory control, storage, and security; wasting infrastructure energy for lights, forklift energy, and possible cooling or heating of the inventory spaces if the goods require temperature or humidity control; wasting space and the associated lease/mortgage fees and taxes; and becoming scrap if not sold (a frequent waste in large inventories). Inventory and inventory space reductions lead to infrastructure energy savings. Process energy can also be saved by not making the FG that end up in inventory, cannot be sold, and become scrap. Refs. 3 and 4 contain case studies for, among others, inventory reductions. A convenient nondimensional metric to track the overall impact of all inventories on energy savings is

$$EC_T \times \frac{I_T}{AC} \tag{2}$$

where I_T is the number of inventory turns per year.

Workmanship, Training, and Quality Assurance

In the ideal Lean system, the processes, equipment, procedures, and training are perfected to the degree that guarantees consistent and robust production with predictable effort, timing, quality, and cost; with no variability, defects, or rework, and with maximum ergonomics and safety. This is accomplished by a consistent long-term strategy of continuous improvement of all the above elements, including intensive initial training of the workforce and subsequent retraining in new procedures. A procedure must be developed for each process until it is robust and predictable and optimized for

Example 1 Energy waste from poor workmanship

A plant with \$20,000,000 in sales and \$2,000,000 in profits spends \$1,000,000 on energy per year. The typical order requires 10 processes of roughly equal energy consumption. The production equipment consumes 60% and the supportive infrastructure consumes 40% of the plant energy. Sequential process #5 has the defect rate of 10%. In order to compensate for the defects, the first 5 processes must produce 10% extra pieces. The annual waste of energy cost (and the energy cost savings, if the defective process is fixed) is then:

$$(\$1,000,000/\text{yr})(5/10 \text{ processes})(60\% \text{ process energy})(10\% \text{ defect rate}) = \$30,000/\text{yr} \quad (3)$$

The additional production time of 10% wastes not only the cost of the process energy computed in (3) but also the infrastructure energy cost of:

$$(\$1,000,000/\text{yr})\ (40\% \text{ infrastructure energy})\ (10\% \text{ defects}) = \$40,000/\text{yr} \quad (4)$$

Such delays also extend the promised delivery time and reduce customer satisfaction and factory competitiveness. Adding (1) and (2) together, (not counting the direct productivity losses), the wasted energy cost alone of \$70,000/yr represents 3.5% of the annual profits and 7% in annual energy costs. Based on the author's experience,[5] these numbers are not infrequent in industry. Fixing the productivity of process #5 would eliminate these wastes.

minimum overall cost, required quality, maximum ergonomics, and safety. Process operators must be trained in the procedures as well as in the process quality assurance, and they must be empowered to stop the process and take corrective action or call for help if unable to avoid a defect. Management culture must be supportive for such activities. Any departure from this ideal leads to costly penalties in quality, rework, delays, overtime or contract penalties, crew frustrations, and customer dissatisfaction. These, in turn, have negative impacts on energy as follows:

1. Defects require rework, which requires additional energy to remake or repair the part. The best metric to use here is the energy or energy cost per part used in the given defective process multiplied by the number of bad parts produced per year.
2. Variability in the process time or delays caused by defects mean that the production takes more time and more infrastructure and process energy for the same amount of value work and profits when compared with the ideal nonvariable process. Example 1 illustrates cases (1) and (2).
3. Defective processes usually require a massive final inspection to sort out the good products. Finding the finished goods defective is the most inefficient means of quality assurance because often the entire batch must then be remade, consuming the associated energy. The inspection space, labor, and energy represent a direct waste and should be replaced with in-line quality assurance (ILQA) that detects the first bad piece (Governmental, medical, etc. orders usually require a 100% final inspection. In the Lean system, this is performed as a formality because everybody in the plant knows that all pieces will be perfect because all imperfections have been removed in real time before the inspection process.) and immediately undertakes a corrective action. Typically, the ILQA can be implemented in few days of operators' training and has the simple payback period measured in days or weeks.[5]

Overage Reduction

Many a plant compensates for its notorious defects by routinely scheduling production in excess of what the customer orders. Some minimum overage is usually justified for machine setups, adjustments, and QA samples. In a Lean plant this rarely exceeds a fraction of one percent. In a traditional plant, the value of 5%–15% is not infrequent. A 5% overage means that the plant spends 105% of the necessary costs. If the profit margin is 5%, the overage alone may consume the entire profit. The overall energy waste (and the opportunity to save energy) is simply proportional to the overage amount. Overage is one of the most wasteful ways of compensating for defective processes. The best remedy is to simply identify the defective process with ILQA, find the root cause (typically the lack of training, excessive work quotas, or bad process or material), and repair it.

Unintentional overage can also be destructive to profits and energy use. Example: A worker is asked to cut only a few small pieces from a large sheet of metal, but instead he cuts the entire sheet, thinking, "my machine is already setup and soon they will ask me to cut the rest of the sheet anyway, so I may as well do it now". The excessive pieces then move through all processes, unknowingly to the management, consuming energy, labor and fixed costs, to end up as excessive FG inventory and, in the worst case, find no buyer and end up as scrap. Uncontrolled and careless overage can easily consume all profits, and, of course, waste energy proportionately to the overage amount.

Example 2 Energy savings from setup time reduction

A plant operates on two shifts, 260 days per year, performing on average of 20 two-hour setups per day on their electrically heated injection molding machines. Each machine consumes 20 kW when idle but energized. By a focused continuous improvement system and training, the crew reduces the routine setup time to 0.5 h, with few, if any expenses for additional hardware, thus saving:

(260 days/yr) (20 setups/day) (1.5 h saved/setup) = 7800 machine h/yr.

The resultant process energy saved will be:

$$(7800\ \text{h/year})\ (20\ \text{kW}) = 156{,}000\ \text{kWh/yr} \quad (5)$$

In addition, infrastructure energy will be saved because of the reduced downtime. Using the data from Example 1, if the work is done in two shifts for 260 days per year (4160 h/yr), the plant infrastructure uses 40% of the plant energy, and each machine consumes 2% of the plant infrastructure energy during the setup, the additional energy cost savings due to the setup time reduction will be:

$$(7{,}800\ \text{hr/yr})\ (0.02)\ (0.04)\ (\$1{,}000{,}000)/(4160\ \text{h/yr}) = \$15{,}000 \quad (6)$$

Downtime

Equipment downtime and idleness may occur due to scheduled maintenance, unscheduled breakdowns, machine setups, and poor process scheduling. The downtime may cause proportional loss of both profits and energy. The downtime may have fourfold impact on energy use, as follows:

1. When a process stops for whatever reason during an active production shift, the plant infrastructure continues to use energy and loosing money, as in Eq. 4. A good plant manager should understand what fraction of the infrastructure energy is wasted during the specific equipment downtime. With this knowledge, the energy waste can be estimated as being proportional to the downtime.
2. Some machines continue using energy during maintenance, repair, or setup in proportion to the downtime (e.g., the crucible holding molten metal for a die casting machine remains heated by natural gas while the machine is being setup or repaired). Reducing the setup time or eliminating the repair time saves the gas energy in direct proportion to the downtime saved. In order to calculate energy savings in such situations, it is necessary to understand the energy consumption by the equipment per unit of time multiplied by the downtime reduction.
3. When a particular machine is down, additional equipment upstream or downstream of that machine may also be forced into an idle status but remain energized, thus wasting energy. In an ideal single-piece flow, the entire production line (As in the saying "In Lean either everything works or nothing works.") will stop. In order to estimate the energy-saving opportunity from reducing this cumulative downtime, the energy manager must understand which equipment is idled by the downtime of a given machine and how much energy it uses per unit time while being idle.
4. Lastly, energized equipment should be well managed. A high-powered machine may be left energized for hours at a time when not scheduled for production. A good practice is to assign each of these machines to an operator who will have the duty of turning the machine off when not needed for a longer time, if practical, and to turn it back on just in time to be ready for production exactly when it is needed.

Preventive maintenance and setup time reduction have a particularly critical impact on both productivity and related energy use, as follows:

Preventive Maintenance

Practical and routine preventive maintenance should be done during the hours free of scheduled production (e.g., during night shifts, on weekends, or during layover periods). The maintenance should be preventive rather than reactive (The term "preventive" tends to be replaced with "productive" in modern industrial parlance). Well-managed "total" preventive[6] maintenance involves not only oiling and checking the machines per schedule but also ongoing training of the mechanics; developing a comprehensive database containing information on the particular use and needs of various machines; preparing a schedule of part replacement and keeping inventory of frequently used spare parts; and a well-managed ordering system for other parts, including vendor data so that when a part is needed it can be ordered immediately and shipped using the fastest possible means. Industry leaders have demonstrated that affordable preventive maintenance can reduce the unscheduled downtime and associated energy waste to zero. This should be the practical goal of well-run factories.

Setups

Modern market trends push industry towards shorter series and smaller orders, requiring, in turn, more and shorter setups. Industry leaders have perfected routine setups to take no more than a few minutes. In poorly managed plants, routine setups can take as long as several hours. In all competitive modern plants, serious efforts should be devoted to setup time reductions. The effort includes both training and hardware improvements. The training alone, with only minimal additional equipment (such as carts), can yield dramatic setup time reductions (i.e., from hours to minutes). Further gains may require a change of the mounting and adjustment hardware and instrumentation. Some companies organize competitions between teams for developing robust procedures for the setup time reductions. In a plant performing many setups, the opportunity for energy savings may be significant, both in the process and infrastructure energy, as shown in Example 2.

Flexibility

Production flexibility, also called agility, is an important characteristic of competitive plants. A flexible plant prefers small machines (if possible, on casters) that are easy to roll into position and plug into adjustable quick-connect electrical and air lines and that are easy to setup and maintain over the large fixed machines selected with large batches and small unit costs in mind (such machines are called "monuments" in Ref. 3). Such an ideal plant will also have trained a flexible workforce in multiple skills, including quality assurance skills. This flexibility allows for the setup of new production lines in hours or even minutes, optimizing the flow and floor layout in response to short orders, and delivers the orders JIT. The energy may be saved in two important ways, as follows:

- Small machines processing one piece at a time use only as much energy as needed. In contrast, when excessively large automated machines are used, the typical management choice is between using small batches JIT, thus wasting the large machine energy, or staging the batches for the large machine, which optimizes machine utilization at the expense of throughput time, production flow, production planning effort, and the related infrastructure energy.
- Small machines are conducive to flexible cellular work layout, where 2–4 machines involved in the sequential processing of WIP are arranged into a U-shaped cell with 1–3 workers serving all processes in the cell in sequence, and the last process being quality assurance. This layout can be made very compact—occupying a much smaller footprint in the plant compared to traditional "process village" plants, roughly a reduction of 50%[3,4]—and is strongly preferred by workers because it saves walking and integrates well the work steps. Such a layout also saves forklift effort and energy and infrastructure energy due to the reduction of the footprint.

Other Productivity Elements

The complete list of productivity elements is beyond the scope of this article, and all elements have some leverage on energy use and conservation. In the remaining space, only the few most important remaining aspects are mentioned, with their leverage on energy. Descriptive details can be found in Ref. 7 and numerous other texts on Lean production.

- *Visual factory*: Modern factories place an increasing importance on making the entire production as transparent as possible in order to make any problem visible to all, which is motivational for immediate corrective actions and continuous improvements. Ideally, each process should have a white board displaying the short-term data, such as the current production status (quantity completed vs required); the rate of defects or rejects and their causes; control charts and information about the machine condition or maintenance needs; and a brief list and explanation of any issues, all frequently updated. The board should also display long-term information such as process capability history, quality trends, operator training, etc. Such information is most helpful in the optimization of, among other things, process time and quality, which leads to energy savings, as discussed above.
- "*Andon*" *signals*: The term refers to the visual signals (lights, flags, markers, etc.) displaying the process condition, as follows: "green = all OK," "yellow = minor problem being corrected," and "red = high alarm, stopped production, and immediate assistance needed." The signals are very useful in identifying the trouble-free and troubled processes, which is conducive to focusing the aid resources to the right places in real time, fixing problems immediately and not allowing defects to flow downstream on the line. These features, in turn, reduce defects, rework, delays, and wasted costs, which improve overall productivity and save energy, as described above. It is also useful to display the estimated downtime (Toyota and other modern plants have large centrally located Andon boards that display the Andon signal, the workstation number, and the estimated downtime.). Knowing the forecasted downtime frees other workers to perform their pending tasks which have waited for such an opportunity rather than wait idle. This leads to better utilization of the plant resources, including infrastructure energy.

Fig. 2 In this messy plant, the workers waste close to 20% of their time looking for items and scavenging for parts and tools, also wasting the plant energy.

- "*5Ss*": The term comes from five Japanese words that begin with the "s" sound and loosely translate into English as: sorting, simplification, sweeping, standardization, and self-discipline (many other translations of the words are popular in industry); and describes a simple but powerful workplace organization method.[8] The underlying principle of the method is that only the items needed for the immediate task (parts, containers, tools, instructions, materials) are kept at hand where they are needed at the moment, and everything else is kept in easily accessible and well-organized storage in perfect order, easy to locate without searching, and in just the right quantities. All items have their designated place, clearly labeled with signs, labels, part numbers, and possibly bar codes. The minimum and maximum levels of inventory of small parts are predefined and are based on actual consumption rather than the "just-in-case" philosophy. The parts, tools, and materials needed for the next shift of production are prepared by a person in charge of the storage during the previous shift and delivered to the workstation before the shift starts. The floor is uncluttered and marked with designated spaces for all equipment. The entire factory is spotlessly clean and uncluttered. Walls are empty except for the visual boards. In consequence of these changes, the searching for parts, tools, and instructions which can represent a significant waste of labor and time is reduced, and this, in turn, saves energy. Secondary effects are also important. In a well-organized place, fewer mistakes are made; fewer wrong parts are used; less inspection is needed; quality, throughput time, and customer satisfaction are increased; and costs and energy are decreased. Fig. 2 illustrates a fragment of a messy factory, where the average worker was estimated to waste 20% of his shift time looking for and scavenging for parts and tools. This percentage multiplied by the number of workers yields a significant amount of wasted production time, also wasting plant energy in the same proportion. Sorting, cleaning, and organizing the workplace is one of the simplest and most powerful starting points on the way to improved productivity and energy savings.

CONCLUSION

Large savings in energy are possible as an inherent byproduct of improving productivity. The state-of-the-art Lean productivity method can yield dramatic improvements in productivity. In the extreme case of converting from the traditional batch-and-queue and "process village" manufacturing system to Lean production,

Em–Energy Con

overall costs, lead times, and inventories can be reduced by as much as 50%–90%, floor space and energy by 50%, and energy density can be improved by 50%. The amount of energy that can be saved by productivity improvements often radically exceeds the savings from equipment optimization alone, thus providing a strong incentive to include productivity improvements in energy-reduction efforts.

Productivity strongly depends on human factors such as management, learning, and training, communications, culture, teamwork, etc. which are difficult to quantify, making accurate estimates of the cost, schedule, and quality benefits from various productivity improvements and the related energy savings difficult to estimate with engineering accuracy. For this reason, simple metrics and models are recommended, and some examples have been presented. If applied conservatively, they can become useful tools for energy management in a plant. The prerequisite knowledge includes an understanding of Lean Flow and its various productivity elements and a good accounting of energy use in the plant, including the knowledge of the energy used by individual machines and processes both when in productive use and in the idle but energized state, as well as the energy elements used by the infrastructure (various light combinations, air-compressors, cooling and heating devices, combusting systems, conveyers, forklifts, etc.). In the times of ferocious global competition and rising energy prices, every industrial plant should make every effort to improve both productivity and energy use.

ACKNOWLEDGMENTS

This work is a result of the studies of energy conservation using the Lean productivity method performed by the Industrial Assessment Center funded by the U.S. Department of Energy at Loyola Marymount University. The author is grateful to Mr. Rudolf Marloth, Assistant Director of the Center, for his help with various energy estimates included herein and his insightful comments, to the Center students for their enthusiastic work, and to his son Peter W. Oppenheim for his diligent editing.

REFERENCES

1. U.S. Department of Energy, Energy Efficiency and Renewable Energy, http://www.eere.energy.gov/industry/ (accessed on December 2005).
2. U.S. Department of Energy, http://eereweb.ee.doe.gov/industry/bestpractices/plant_assessments.html, (accessed on December 2005).
3. Womack, P.J.; Jones, D.T. *Lean Thinking. 2nd Ed.*; Lean Enterprise Institute: Boston, MA, 2005; (www.lean.org), ISBN: 0-7432-4927-5.
4. Liker, J. *Becoming Lean, Inside Stories of U.S. Manufacturers*; Productivity Inc.: Portland, OR, 1998; service@productivityinc.com.
5. Oppenheim, B.W. *Selected Assessment Recommendations*, Industrial Assessment Center, Loyola Marymount University: Los Angeles, (boppenheim@lmu.edu), unpublished 2004.
6. *Setup Reduction for Just-in-Time*, Video/CD, Society of Manufacturing Engineers, Product ID: VT90PUB2, http://www.sme.org/cgi-bin/get-item.pl?VT392&2&SME&1990, (accessed on December 2005).
7. Ohno, T. *Toyota Production System: Beyond Large Scale Production*; Productivity Press: New York, 1988; info@productivityinc.com.
8. Hiroyuki, H. *Five Pillars of the Visual Workplace, the Sourcebook for 5S Implementation*; Productivity Press: New York, NY, 1995; info@productivityinc.com.

Energy Conversion: Principles for Coal, Animal Waste, and Biomass Fuels

Kalyan Annamalai
Department of Mechanical Engineering, College Station, Texas, U.S.A.

Soyuz Priyadarsan
Texas A&M University, College Station, Texas, U.S.A.

Senthil Arumugam
Enerquip, Inc., Medford, Wisconsin, U.S.A.

John M. Sweeten
Texas Agricultural Experiment Station, Amarillo, Texas, U.S.A.

Abstract

A brief overview is presented of various energy units; terminology; and basic concepts in energy conversion including pyrolysis, gasification, ignition, and combustion. Detailed sets of fuel properties of coal, agricultural biomass, and animal waste are presented so that their suitability as a fuel for the energy conversion process can be determined. It is also found that the dry ash free (DAF) heat values of various biomass fuels, including animal waste, remain approximately constant, which leads to a presentation of generalized results for maximum flame temperature as a function of ash and moisture contents. The cofiring technology is emerging as a cost-effective method of firing a smaller percentage of biomass fuels, with coal as the major fuel. Various techniques of cofiring are summarized. Gasification approaches, including FutureGen and reburn technologies for reduction of pollutants, are also briefly reviewed.

INTRODUCTION AND OBJECTIVES

The overall objective of this entry is to provide the basics of energy conversion processes and to present thermochemical data for coal and biomass fuels. Energy represents the capacity for doing work. It can be converted from one form to another as long as the total energy remains the same. Common fuels like natural gas, gasoline, and coal possess energy as chemical energy (or bond energy) between atoms in molecules. In a reaction of the carbon and hydrogen in the fuel with oxygen, called an oxidation reaction (or more commonly called combustion), carbon dioxide (CO_2) and water (H_2O) are produced, releasing energy as heat measured in units of kJ or Btu (see Table 1 for energy units). Combustion processes are used to deliver (i) work, using external combustion (EC) systems by generating hot gases and producing steam to drive electric generators as in coal fired power plans, or internal combustion (IC) engines by using the hot gases directly as in automobiles or gas turbines; and (ii) thermal energy, for applications to manufacturing processes in metallurgical and chemical industries or agricultural product processing.

Fuels can be naturally occurring (e.g., fossil fuels such as coal, oil, and gas, which are residues of ancient plant or animal deposits) or synthesized (e.g., synthetic fuels). Fuels are classified according to the phase or state in which they exist: as gaseous (e.g., natural gas), liquid (e.g., gasoline or ethanol), or solid (e.g., coal, wood, or plant residues). Gaseous fuels are used mainly in residential applications (such as water heaters, home heating, or kitchen ranges), in industrial furnaces, and in boilers. Liquid fuels are used in gas turbines, automotive engines, and oil burners. Solid fuels are used mainly in boilers and steelmaking furnaces.

During combustion of fossil fuels, nitrogen or sulfur in the fuel is released as NO, NO_2 (termed generally as NO_x) and SO_2 or SO_3 (termed as SO_x). They lead to acid rain (when SO_x or NO_x combine with H_2O and fall as rain) and ozone depletion. In addition, greenhouse gas emissions (CO_2, CH_4, N_2O, CFCs, SF_6, etc.) are becoming a global concern due to warming of the atmosphere, as shown in Fig. 1 for CO_2 emissions. Global surface temperature has increased by 0.6°C over the past 100 years. About 30%–40% of the world's CO_2 is from fossil fuels. The Kyoto protocol, signed by countries that account for 54% of the world's fossil based CO_2 emissions, calls for reduction of greenhouse gases by 5% from 1990 levels over the period from 2008 to 2012.

The total worldwide energy consumption is 421.5 quads of energy in 2003 and is projected to be 600 quads in 2020, while U.S. consumption in 2004 is about 100 quads and is projected to be 126 quads in 2020. The split

Keywords: Energy units; Fuels; Biomass; Pyrolysis; Gasification; Cofiring; Reburn; Heating value; Pollutants.

Table 1 Energy units and terminology

The section on energy Units and Conversion factors in Energy is condensed from Chapter 01 of Combustion Engineering by Annamalai and Puri [2005] and Tables.

Energy Units

1 Btu (British thermal unit) = 778.14 ft lb_f = 1.0551 kJ, 1 kJ = 0.94782 Btu = 25,037 $lbmft/s^2$

1 mBtu = 1 k Btu = 1000 Btu, 1 mmBtu = 1000 k Btu = 10^6 Btu, 1 trillion Btu = 10^9 Btu or 1 giga Btu

1 quad = 10^{15} Btu or 1.05×10^{15} kJ or 2.93×10^{11} kW h,

1 Peta J = 10^{15} J = 10^{12} kJ»0.00095 Quads

1 kilowatt-hour of electricity = 3,412 BTU = 3.6 Mj,

1 cal: 4.1868 J, One (food) calorie = 1000 cal or 1 Cal,

1 kJ/kg = 0.43 Btu/lb, 1 Btu/lb = 2.326 kJ/kg

1 kg/GJ = 1 g/MJ = 2.326 lb/mmBtu; 1 lb/mmBtu = 0.430 kg/GJ = 0.430 g/MJ

1 Btu/SCF = 37 kJ/m^3

1 Therm = 10^5 Btu = 1.055×10^5 kJ

1 m^3/GJ = 37.2596 ft^3/mmBTU

1 hp = 0.7064 Btu/s = 0.7457 kW = 745.7 W = 550 lbf ft/s = 42.41 Btu/min

1 boiler HP = 33475 Btu/h, 1 Btu/h = 1.0551 kJ/h

1 barrel (42 gal) of crude oil = 5,800,000 Btu = 6120 MJ

1 gal of gasoline = 124,000 Btu = 131 MJ

1 gal of heating oil = 139,000 Btu = 146.7 MJ, 1 gal of diesel fuel = 139,000 Btu = 146.7 MJ

1 barrel of residual fuel oil = 6,287,000 Btu = 6633 MJ

1 cubic foot of natural gas = 1,026 Btu = 1.082 MJ, 1 Ton of Trash = 150 kWh

1 gal of propane = 91,000 Btu = 96 MJ, 1 short ton of coal = 20,681,000 Btu = 21821 MJ

Emission reporting for pollutants: (i) parts per million (ppm), (ii) normalized ppm, (iii) emission Index (EI) in g/kg fuel, (iv) g/GJ, v) mg/m^3 of flue gas:

Conversions in emissions reporting: (ii) normalized ppm = ppm × (21-O_2% std)/(21-O_2% measured); (iii) EI of species k: C % by mass in fuel × mol Wt of *k* × ppm of species $k \times 10^{-3}$/{12.01(CO_2% + CO%)}, (iv) g/GJ = EI/ {HHV in GJ/kg}; (v) mg/m^3 = ppm of species *k* × Mol Wt of *k*/24.5

Volume of 1 kmol (SI) and 1 lb mole (English) of an ideal gas at STP conditions defined below:
Pressure at 101.3 kPa (1 atm, 14.7 psia, 29.92 in.Hg, 760 Torr) fixed; *T* changes depending upon type of standard adopted

Scientific (or SATP, standard ambient *T* and *P*)	US standard (1976) or ISA (International standard atmosphere)	NTP (gas industry reference base)	Chemists-standard-atmosphere (CSA)
25°C (77°F)	15°C (60°F)	20°C(68°F), 101.3 kPa	0°C (32°F),
24.5 m^3/kmol (392 ft^3/lb mole); $\rho_{air,SATP}$ = 1.188 kg/m^3 = 0.0698 lb_m/ft^3	23.7 m^3/kmol (375.6 ft^3/lb mole); $\rho_{air,ISA}$ = 1.229 kg/m^3 = 0.0767 lb_m/ft^3	24.06 m^3/kmole or 385 ft^3/lb mole; $\rho_{air,NTP}$ = 1.208 kg/m^3 = 0.0754 lb_m/ft^3	22.4 m^3/kmol (359.2 ft^3/lb mole), $\rho_{air,CSA}$ = 1.297 kg/m^3 = 0.0810 lb_m/ft^3

is as follows: 40 quads for petroleum, 23 for natural gas, 23 for coal, 8 for nuclear power, and 6 for renewables (where energy is renewed or replaced using natural processes) and others sources. Currently, the United States relies on fossil fuels for 85% of its energy needs. Soon, the U.S. energy consumption rate which distributed as electrical power (40%), transportation (30%), and heat (30%), will outpace the growth in the energy production rate, increasing reliance on imported oil. The Hubbert peak theory (named after Marion King Hubbert, a geophysicist with Shell Research Lab in Houston, Texas) is based upon the rate of extraction and depletion of conventional fossil fuels, and predicts that fossil-based oil would peak at about 12.5 billion barrels per year worldwide some time around 2000. The power cost and percentage use of coal in various U.S. states varies from

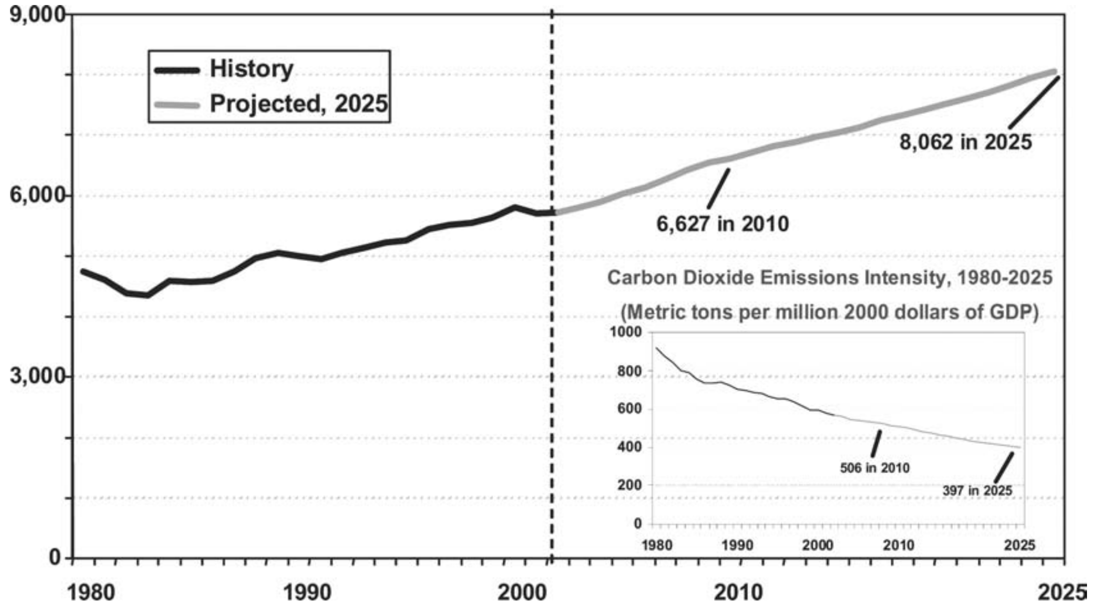


Fig. 1 Total CO_2 emission in million metric tons per year: History and Projected 1980–2025. Source: From DOE-EIA (see Ref. 1).

10 cents (price per kWh) at 1% coal use for power generation in California to 48 cents at 94% use of coal in Utah.

Biomass is defined as "any organic material from living organisms that contains stored sunlight (solar energy) in the form of chemical energy."[1] These include agro-based materials (vegetation, trees, and plants); industrial wastes (sawdust, wood chips, and crop residues); municipal solid wastes (MSWs), which contain more than 70% biomass (including landfill gases, containing almost 50% CH_4); and animal waste. Biomass is a solid fuel in which hydrogen is locked with carbon atoms. Biomass production worldwide is 145 billion metric tons. Biomass now supplies 3% of U.S. energy, and it could be increased to as high as 20%. Renewable energy sources (RES) include biomass, wind, hydro, solar, flowing water or hydropower, anaerobic digestion, ocean thermal (20°C temperature difference), tidal energies, and geothermal (a nonsolar source of energy), and these supply 14% of the world demand. The RES constitute only 6%, while coal, petroleum, and natural gas account for 23%, 40%, and 24%, respectively. About 9% of the world's electricity is from RES, and 65% of the electricity contributed by biomass. About 97% of energy conversion from biomass is by combustion. Many U.S. states have encouraged the use of renewables by offering REC (Renewable Energy Credits). One REC = 1 MW/h = 3.412 mmBtu; hence the use of 1 REC is equivalent to replacing approximately 1500 lb of coal, reducing emission of NO_x and SO_x by 1.5 lb for every 1 REC, assuming that emissions of NO_x and SO_x are 0.45 lb per mmBtu generated by coal. Several emission-reporting methods and conversions are summarized in Table 1. Recently, H_2 is being promoted as a clean-burning, non-global-warming, and pollution-free fuel for both power generation and transportation.

Fig. 2 shows a comparison between biomass and hydrogen energy cycles. In the biomass cycle, photosynthesis is used to split CO_2 into C and O_2, and H_2O into H_2 and O_2, producing Hydrocarbons (HC) fuel (e.g., leaves) and releasing O_2. The O_2 released is used to combust the HC and produce CO_2 and H_2O, which are returned to produce plant biomass (e.g., leaves) and O_2. On the other hand, in the hydrogen cycle, H_2O is disassociated using the photo-splitting process to produce H_2 and O_2, which are then used for the combustion process. The hydrogen fuel can be used in fuel cells to obtain an efficient conversion. Photosynthesis is water intensive; most of the water supplied to plants evaporates through leaves into the atmosphere, where it re-enters the hydrology cycle.

This entry is organized in the following format: (i) coal and biosolid properties; (ii) coal and biosolid pyrolysis (a process of thermal decomposition in the absence of oxygen), combustion, and gasification; (iii) combustion by cofiring coal with biosolids; (iv) gasification of coal and biosolids (a process that includes pyrolysis, partial oxidation due to the presence of oxygen, and hydrogenation); and (v) reburn for NO_x reduction.

FUEL PROPERTIES

Fuel properties play a major role in the selection, design, and operation of energy conversion systems.

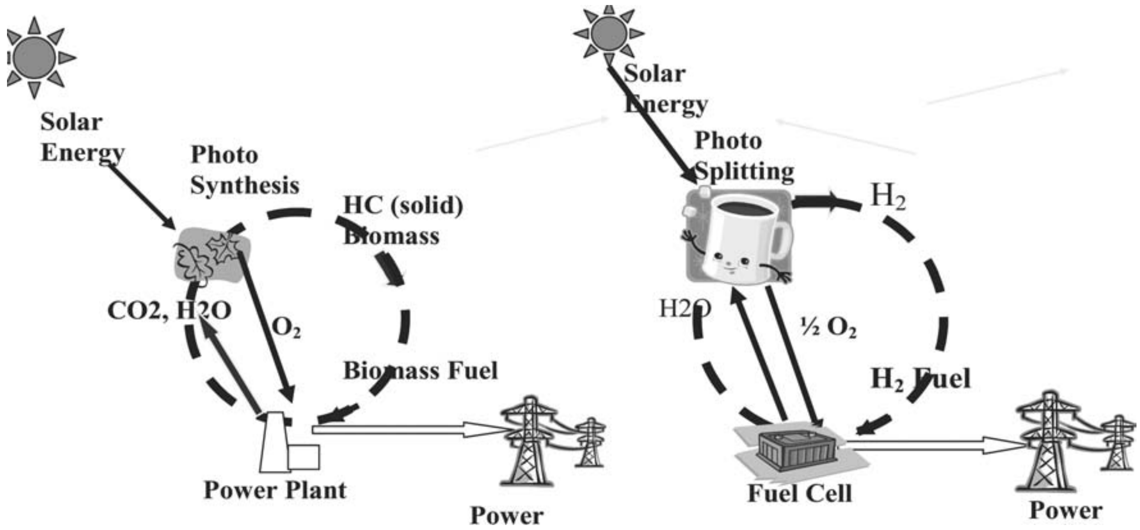


Fig. 2 Comparison of biomass energy and H_2 energy cycles.

Solid Fuels

The primary solid fuel widely used in power plants is coal containing combustibles, moisture, and intrinsic mineral matter originating from dissolved salts in water. During the "coalification" process, lignite, the lowest rank of coal (low C/O ratio), is produced first from peat, followed by sub-bituminous (black lignite, typically low sulfur, noncaking), bituminous (soft coal that tends to stick when heated and is typically high in S), and finally anthracite (dense coal; has the highest carbon content, >90%, low volatile <15%) with a gradual increase in the coal C/O ratio. The older the coal, the higher its rank. Anthracite (almost carbon) is the highest-ranked coal, with a high heating value. To classify coals and ascertain the quality of coal, it is essential to perform proximate and ultimate analyses according to American Society of Testing Materials (ASTM) standards.

Proximate Analysis (ASTM D3172)

A solid fuel consists of combustibles, ash, and moisture. Combustibles together with ash are called the solid content of fuel. A proximate analysis provides the following information: surface moisture (SM) or dry loss (DL), i.e., moisture in air-dried coal; the inherent moisture in the coal (M); volatile matter (VM; produced by pyrolysis, a thermal decomposition process resulting in release of water, gases, oil and tar); fixed carbon (FC; skeletal matter left after release of volatiles); mineral matter (MM; inert collected with solid fuel); and heating value (HV). On combustion, the MM may be partially oxidized or reduced, and the material left after combustion of C and H in the fuel is called ash (CaO, $CaCO_3$, Fe_2O_3, FeO, etc).

Table 2 shows comparative proximate analyses of coal, advanced feedlot biomass (FB, low-ash cattle manure; see "Coal and Bio-Solids Cofiring"), and litter biomass (LB, chicken manure).[2] Feedlot manure has higher moisture, nitrogen, chlorine, and ash content than coal. With aging or composting, the VM in manure decreases as a result of the gradual release of hydrocarbon gases or dehydrogenation, but fuel becomes more homogeneous.

Ultimate/Elemental Analysis (ASTM D3176)

Ultimate analysis is used to determine the chemical composition of fuels in terms of either the mass percent of their various elements or the number of atoms of each element. The elements of interest are C, H, N, O, S, Cl, P, and others. It can be expressed on an "as received" basis, on a dry basis (with the moisture in the solid fuel removed), or on a dry ash free (DAF) basis (also known as the moisture ash free basis MAF). Tables 3 and 4 show the ultimate analyses of various types of coal and biomass fuels.[3] While nitrogen is not normally present in natural gas, coal has 1%–1.5%; cattle manure and chicken waste contain high amounts of N (Table 2).

Heating Value (ASTM D3286)

The gross or higher heating value (HHV) of a fuel is the amount of heat released when a unit (mass or volume) of the fuel is burned. The HHV of solid fuel is determined using ASTM D3286 with an isothermal jacket bomb calorimeter. For rations fed to animals and animal waste fuels, the HHV for DAF roughly remains constant at about 19,500 kJ/kg (8400 Btu/lb),[4] irrespective of stage of decomposition of animal waste. The HHV can also be

Em—Energy Con

Table 2 Coal, advanced feedlot biomass (FB) and litter biomass (LB)

Parameter	Wyoming coal	Cattle manure (FB)	Chicken manure (LB)[a]	Advanced Feedlot biomass (AFB)[b]	High-ash Feedlot biomass (HFB)[b]
Dry loss (DL)	22.8	6.8	7.5	10.88	7.57
Ash	5.4	42.3	43.8	14.83	43.88
FC	37.25	40.4	8.4	17.33	10.28
VM	34.5	10.5	40.3	56.97	38.2
C	54.1	23.9	39.1	50.08	49.27
H	3.4	3.6	6.7	5.98	6.13
N	0.81	2.3	4.7	38.49	38.7
O	13.1	20.3	48.3	4.58	4.76
S	0.39	0.9	1.2	0.87	0.99
Cl	<0.01%	1.2			
HHV-as received (kJ/kg)	21385	9560	9250	14983	9353
$T_{adiab, Equil}$[c]	2200 K (3500°F)	2012 K (3161°F)			
DAF formula	$CH_{0.76}O_{0.18}N_{0.013}S_{0.0027}$	$CH_{1.78}O_{.64}N_{.083}S_{.014}$	$CH_{2.04}O_{0.93}N_{0.10}S_{0..012}$	$CH_{1.4184}O_{0.5764}N_{0.078}S_{0.0066}$	$CH_{1.4775}O_{0.5892}N_{0.083}S_{0.0076}$
HHV–DAF (kJ/kg)	29785	18785	18995	20168	19265
CO_2, g/GJ					
N, g/GJ					
S, g/GJ					

[a] Ref. 2.

[b] Ref. 37.

[c] Equilibrium temperature for stoichiometric mixture from THERMOLAB Spreadsheet software for any given fuel of known composition (Ref. 36. website http://www.crcpress.com/e_products/downloads/download.asp?cat_no=2553)

Table 3 Coal composition (DAF basis)

ASTM Rank	State (U.S.A.)	Ash, % (dry)	C	H	N	S*	O**	HHV_{Est} kJ/kg	CO_2 kg/GJ	N kg/GJ	S kg/GJ
Lignite	ND	11.6	63.3	4.7	0.48	0.98	30.5	24,469	94.8	0.196	0.401
Lignite	MT	7.7	70.7	4.9	0.8	4.9	22.3	28,643	90.4	0.279	1.711
Lignite	ND	8.2	71.2	5.3	0.56	0.46	22.5	28,782	90.7	0.195	0.160
Lignite	TX	9.4	71.7	5.2	1.3	0.72	21.1	29,070	90.4	0.447	0.248
Lignite	TX	10.3	74.3	5	0.37	0.51	19.8	29,816	91.3	0.124	0.171
Sbb. A	WY	8.4	74.3	5.8	1.2	1.1	17.7	31,092	87.6	0.386	0.354
Sbb. C	WY	6.1	74.8	5.1	0.89	0.3	18.9	30,218	90.7	0.295	0.099
HVB	IL	10.8	77.3	5.6	1.1	2.3	13.6	32,489	87.2	0.339	0.708
HVC	IL	10.1	78.8	5.8	1.6	1.8	12.1	33,394	86.5	0.479	0.539
HVB	IL	11.8	80.1	5.5	1.1	2.3	11.1	33,634	87.3	0.327	0.684
HVB	UT	4.8	80.4	6.1	1.3	0.38	11.9	34,160	86.2	0.381	0.111
HVA	WV	7.6	82.3	5.7	1.4	1.8	8.9	34,851	86.5	0.402	0.516
HVA	KY	2.1	83.8	5.8	1.6	0.66	8.2	35,465	86.6	0.451	0.186
MV	AL	7.1	87	4.8	1.5	0.81	5.9	35,693	89.3	0.420	0.227
LV	PA	9.8	88.2	4.8	1.2	0.62	5.2	36,153	89.4	0.332	0.171
Anthracite	PA	7.8	91.9	2.6	0.78	0.54	4.2	34,974	96.3	0.223	0.154
Anthracite	PA	4.3	93.5	2.7	0.24	0.64	2.9	35,773	95.8	0.067	0.179

HHV_{est}: Boie Equation. CO_2 in g/MJ or kg/GJ = C content in % × 36645/{HHV in kJ/kg}. CO_2 in lb per mmBtu = Multiply CO_2 in (g /MJ) or kg/GJ by 2.32. N in g/MJ or kg/GJ = N% × 10000/{HHV in kJ/kg}.
For NO_x estimation, multiply N content in g/MJ by 1.15 to get NO_x in g/MJ which assumes 35% conversion of fuel N.
For SO_2 estimation, multiply S content in g/MJ by 2 to get SO_2 in g/MJ assuming 100% conversion of fuel S (Multiply HHV in kJ/kg by 0.430 to get Btu/lb).
*Organic sulfur; **by difference.

Table 4 Ultimate analyses and heating values of biomass fuels

Biomass	C	H	O	N	S	Residue	Measured HHV_M	[a]Estimated HHV	CO2 g/MJ	N, g/MJ	S, g/MJ
Field crops											
Alfalfa seed straw	46.76	5.40	40.72	1.00	0.02	6.07	18.45	18.27	92.9	0.542	0.011
Bean straw	42.97	5.59	44.93	0.83	0.01	5.54	17.46	16.68	90.2	0.475	0.006
Corn cobs	46.58	5.87	45.46	0.47	0.01	1.40	18.77	18.19	90.9	0.250	0.005
Corn stover	43.65	5.56	43.31	0.61	0.01	6.26	17.65	17.05	90.6	0.346	0.006
Cotton stalks	39.47	5.07	39.14	1.20	0.02	15.10	15.83	15.51	91.4	0.758	0.013
Rice straw (fall)	41.78	4.63	36.57	0.70	0.08	15.90	16.28	16.07	94.0	0.430	0.049
Rice straw (weathered)	34.60	3.93	35.38	0.93	0.16	25.00	14.56	12.89	87.1	0.639	0.110
Wheat straw	43.20	5.00	39.40	0.61	0.11	11.40	17.51	16.68	90.4	0.348	0.063
Switchgrass[b]	42.02	6.30	46.10	0.77	0.18	4.61	15.99	15.97	96.3	0.482	0.113
Orchard prunings											
Almond prunings	51.30	5.29	40.90	0.66	0.01	1.80	20.01	19.69	93.9	0.330	0.005
Black Walnut	49.80	5.82	43.25	0.22	0.01	0.85	19.83	19.50	92.0	0.111	0.005
English Walnut	49.72	5.63	43.14	0.37	0.01	1.07	19.63	19.27	92.8	0.188	0.005
Vineyard prunings											
Cabernet Sauvignon	46.59	5.85	43.90	0.83	0.04	2.71	19.03	18.37	89.7	0.436	0.021
Chenin Blanc	48.02	5.89	41.93	0.86	0.07	3.13	19.13	19.14	92.0	0.450	0.037
Pinot Noir	47.14	5.82	43.03	0.86	0.01	3.01	19.05	18.62	90.7	0.451	0.005
Thompson seedless	47.35	5.77	43.32	0.77	0.01	2.71	19.35	18.60	89.7	0.398	0.005
Tokay	47.77	5.82	42.63	0.75	0.03	2.93	19.31	18.88	90.7	0.388	0.016
Energy Crops											
Eucalyptus											
Camaldulensis	49.00	5.87	43.97	0.30	0.01	0.72	19.42	19.19	92.5	0.154	0.005
Globulus	48.18	5.92	44.18	0.39	0.01	1.12	19.23	18.95	91.8	0.203	0.005
Grandis	48.33	5.89	45.13	0.15	0.01	0.41	19.35	18.84	91.5	0.078	0.005
Casuarina	48.61	5.83	43.36	0.59	0.02	1.43	19.44	19.10	91.6	0.303	0.010
Cattails	42.99	5.25	42.47	0.74	0.04	8.13	17.81	16.56	88.5	0.415	0.022
Popular	48.45	5.85	43.69	0.47	0.01	1.43	19.38	19.02	91.6	0.243	0.005
Sudan grass	44.58	5.35	39.18	1.21	0.08	9.47	17.39	17.63	93.9	0.696	0.046

					Forest residue						
Black Locust	50.73	5.71	41.93	0.57	0.01	0.97	19.71	19.86	94.3	0.289	0.005
Chaparral	46.9	5.08	40.17	0.54	0.03	7.26	18.61	17.98	92.3	0.290	0.016
Madrone	48	5.96	44.95	0.06	0.02	1	19.41	18.82	90.6	0.031	0.010
Manzanita	48.18	5.94	44.68	0.17	0.02	1	19.3	18.9	91.5	0.088	0.010
Ponderosa Pine	49.25	5.99	44.36	0.06	0.03	0.3	20.02	1937	90.1	0.030	0.015
Ten Oak	47.81	5.93	44.12	0.12	0.01	2	18.93	18.82	92.6	0.063	0.005
Redwood	50.64	5.98	42.88	0.05	0.03	0.4	20.72	20.01	89.6	0.024	0.014
White Fur	49	5.98	44.75	0.05	0.01	0.2	19.95	19.22	90.0	0.025	0.005
					Food and fiber processing wastes						
Almond hulls	45.79	5.36	40.6	0.96	0.01	7.2	18.22	17.89	92.1	0.527	0.005
Almond shells	44.98	5.97	42.27	1.16	0.02	5.6	19.38	18.14	85.0	0.599	0.010
Babassu husks	50.31	5.37	42.29	0.26	0.04	1.73	19.92	19.26	92.5	0.131	0.020
Sugarcane bagasse	44.8	5.35	39.55	0.38	0.01	9.79	17.33	17.61	94.7	0.219	0.006
Coconut fiber dust	50.29	5.05	39.63	0.45	0.16	4.14	20.05	19.2	91.9	0.224	0.080
Cocoa hulls	48.23	5.23	33.09	2.98	0.12	10.25	19.04	19.56	92.8	1.565	0.063
Cotton gin trash	39.59	5.26	36.33	2.09		16.68	16.42	16.13	88.4	1.273	0.000
Macadamia shells	54.41	4.99	39.69	0.36	0.01	0.56	21.01	20.55	94.9	0.171	0.005
Olive pits	48.81	6.23	43.48	0.36	0.02	1.1	21.39	19.61	83.6	0.168	0.009
Peach pits	53	5.9	39.14	0.32	0.05	1.59	20.82	21.18	93.3	0.154	0.024
Peanut hulls	45.77	5.46	39.56	1.63	0.12	7.46	18.64	18.82	90.0	0.874	0.064
Pistachio shells	48.79	5.91	43.41	0.56	0.01	1.28	19.26	19.25	92.8	0.291	0.005
Rice hulls	40.96	4.3	35.86	0.4	0.02	18.34	16.14	15.45	93.0	0.248	0.012
Walnut shells	49.98	5.71	43.35	0.21	0.01	0.71	20.18	19.45	90.8	0.104	0.005
Wheat dust	41.38	5.1	35.19	3.04	0.19	15.1	16.2	16.78	93.6	1.877	0.117

[a] HHV based on Boie equation.

[b] Ref. 20; [Adapted from Refs. 3 and 17] See foot note of Table 3.2 for conversions to English units and estimation of NO_x and SO_2 emissions.

estimated using the ultimate analysis of the fuel and the following empirical relation from Boie[5]:

$$\mathrm{HHV_{fuel}(kJ/kg\ \ fuel)} = 35,160\ Y_C + 116,225\ Y_H - 11,090\ Y_O + 6280\ Y_N + 10465\ Y_S \quad (1)$$

$$\mathrm{HHV_{fuel}(BTU/lb\ \ fuel)} = 15,199\ Y_C + 49,965\ Y_H - 4768\ Y_O + 2700\ Y_N + 4499\ Y_S, \quad (2)$$

where Y denotes the mass fraction of an element C, H, O, N, or S in the fuel. The higher the oxygen content, the lower the HV, as seen in biomass fuels.

Annamalai et al. used the Boie equation for 62 kinds of biosolids with good agreement.[6] For most biomass fuels and alcohols, the HHV in kilojoules per unit mass of stoichiometric oxygen is constant at 14,360–14,730 kJ/kg of O_2 (6165–6320 Btu/lb of O_2).[7]

Estimate of CO_2 Emission

Using the Boie-based HVs for any fuel of known elemental composition, one can plot the CO_2 emission in g/MJ (Fig. 3) as a function of H/C and O/C ratios.[8] Comparisons for selected fuels with known experimental HVs are also shown in the same figure. Coal, with H/C ratio ≈ 0.5, releases the highest CO_2, while natural gas (mainly CH_4) emits the lowest CO_2. Because the United States uses fossil fuels for 86% of its energy needs (100 quads), the estimated CO_2 emission is 6350 million ton/year, assuming that the average CO_2 emission from fossil fuels is 70 kg/GJ (methane: 50 kg/GJ vs coal: 90 kg/GJ). Fig. 1 seems to confirm such estimation within a 10% error.

Flame Temperature

Fig. 4 shows a plot of maximum possible flame temperature vs moisture percentage with combustion for biomass fuels. The result can be correlated as follows[4]:

$$T(\mathrm{K}) = 2290 - 1.89\ \mathrm{H_2O} + 5.06\ \mathrm{Ash} - 0.309\ \mathrm{H_2O\ Ash} - 0.180\ \mathrm{H_2O_2} - 0.108\ \mathrm{Ash}^2 \quad (3)$$

$$T(^\circ\mathrm{F}) = 3650 - 3.40\ \mathrm{H_2O} + 9.10\ \mathrm{Ash} - 0.556\ \mathrm{H_2O\ Ash} - 0.324\ \mathrm{H_2O_2} - 0.194\ \mathrm{Ash}^2 \quad (4)$$

The adiabatic flame temperature decreases if the ash and moisture contents increase.

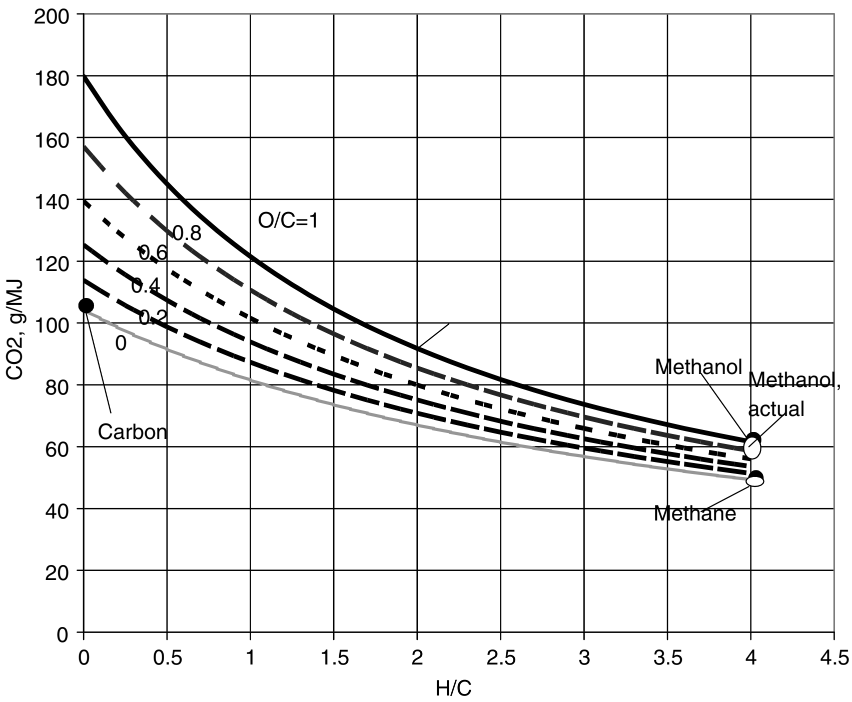


Fig. 3 Emission of CO_2 as a function of H/C and O/C atom ratios in hydrocarbon fuels.
Source: Adapted from Taylor and Francis (see Ref. 8).

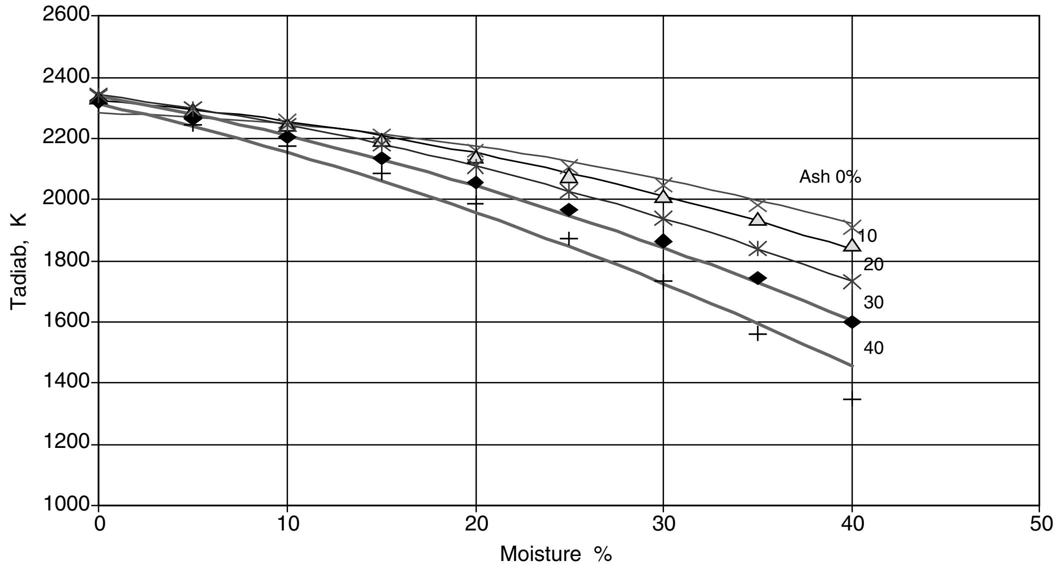


Fig. 4 Correlation of adiabatic flame temperature with moisture and ash contents.

Flue Gas Volume

The flue gas volume for C–H–O is almost independent of O/C ratios. The fit at 6% O_2 in products gives the following empirical equation for flue gas volume (m^3/GJ) at SATP[8]:

$$\text{Flue gas}_{vol}(m^3/GJ) = 4.96\left(\frac{H}{C}\right)^2 - 38.628\left(\frac{H}{C}\right) + 389.72 \quad (5)$$

$$\text{Flue gas}_{vol}(ft^3/mmBtu) = 184.68\left(\frac{H}{C}\right)^2 - 1439.28\left(\frac{H}{C}\right) + 14520.96 \quad (6)$$

Liquid Fuels

Liquid fuels, used mainly in the transportation sector, are derived from crude oil, which occurs naturally as a free-flowing liquid with a density of $\rho \approx 780\ kg/m^3$–$1000\ kg/m^3$, containing 0.1% ash and 0.15%–0.5% nitrogen. Crude oil normally contains a mixture of hydrocarbons, and as such, the "boiling" temperature keeps increasing as the oil is distilled. Most fuel oils contain 83%–88% carbon and 6%–12% (by mass) hydrogen.

Gaseous Fuels

The gaseous fuels are cleaner-burning fuels than liquid and solid fuels. They are a mixture of HC but dominated by highly volatile CH_4 with very little S and N. Natural gas is transported as liquefied natural gas (LNG) and compressed natural gas (CNG), typically at 150–250 bars. Liquefied petroleum gas (LPG) is a byproduct of petroleum refining, and it consists mainly of 90% propane. A low-Btu gas contains 0–7400 kJ/SCM (Standard Cubic Meter, 0–200 Btu/SCF, standards defined in Table 1); a medium-Btu gas. 7400–14,800 kJ/SCM (200–400 Btu/SCF); and a high-Btu gas, above 14,800 kJ/SCM (more than 400 Btu/SCF). Hydrogen is another gaseous fuel, with a heat value of 11,525 kJ/SCM (310 Btu/SCF). Because the fuel quality (heat value) may change when fuel is switched, the thermal output rate at a fixed gas-line pressure changes when fuels are changed.

COAL AND BIOMASS PYROLYSIS, GASIFICATION, AND COMBUSTION

Typically, coal densities range from 1100 kg/m^3 for low-rank coals to 2330 kg/m^3 for high-density pyrolytic graphite, while for biomass, density ranges from 100 kg/m^3 for straw to 500 kg/m^3 for forest wood.[9] The bulk density of cattle FB as harvested is 737 kg/m^3 (CF) for high ash (HA-FB) and 32 lbs/CF for low ash (LA-FB).[10] The processes during heating and combustion of coal are illustrated in Fig. 5, and they are similar for biomass except for high VM. The process of release of gases from solid fuels in the absence of oxygen is called pyrolysis, while the combined process of pyrolysis and

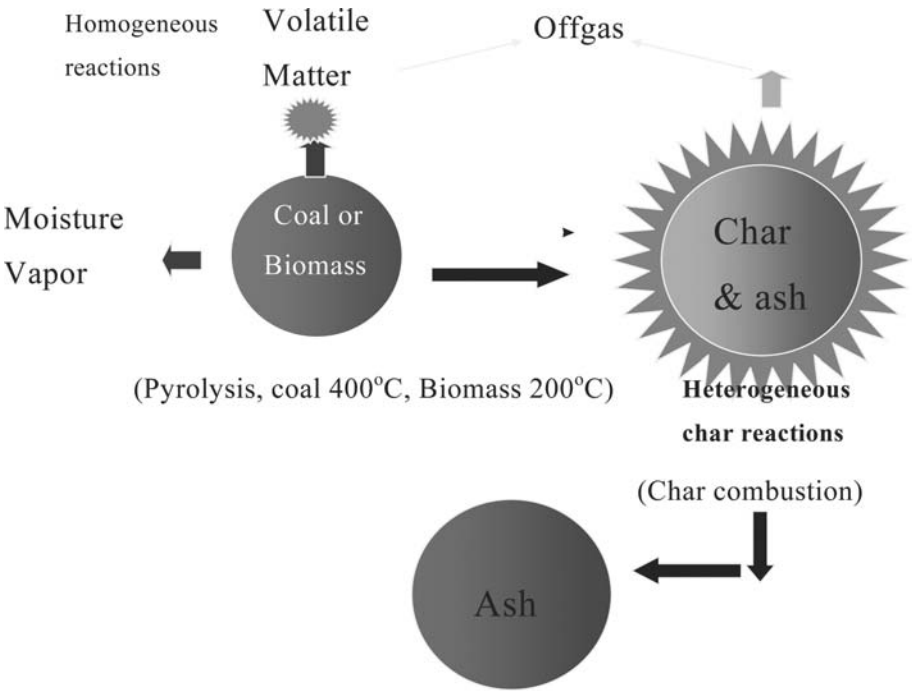


Fig. 5 Processes during coal pyrolysis, gasification, and combustion.

partial oxidation of fuel in the presence of oxygen is known as gasification. If all combustible gases and solid carbon are oxidized to CO_2 and H_2O, the process is known as combustion.

Pyrolysis

Solid fuels, like coal and biomass, can be pyrolyzed (thermally decomposed) in inert atmospheres to yield combustible gases or VM. While biomass typically releases about 70%–80% of its mass as VM (mainly from cellulose and hemicellulose) with the remainder being char, mainly from lignin content of biomass, coal releases 10%–50% of its mass as VM, depending upon its age or rank. Typically, a medium-rank coal consists of 40% VM and 60% FC, while a high-rank coal has about 10% VM. Bituminous coal pyrolyzes at about 700 K (with 1% mass loss for heating rates <100°C/s), as in the case of most plastics. Pyrolytic products range from lighter volatiles like CH_4, C_2H_4, C_2H_6, CO, CO_2, H_2, and H_2O to heavier molecular mass tars. Apart from volatiles, nitrogen is also evolved from the fuel during pyrolysis in the form of HCN, NH_3, and other compounds or, more generally, XN.

Sweeten et al. performed the thermogravimetric analysis (TGA) of feedlot manure.[4] The results are shown in Fig. 6. In the case of manure, drying occurred between 50 and 100°C, pyrolysis was initiated around 185°C–200°C for a heating rate of 80°C/min, and the minimum ignition temperature was approximately 528°C. The gases produced during biomass pyrolysis can also be converted into transportation fuels like biodiesel, methanol, and ethanol, which may be used either alone or blended with gasoline.

Volatile Oxidation

Once released, volatiles (HC, CO, H_2, etc.) undergo oxidation within a thin gas film surrounding the solid fuel particle. The oxidation for each HC involves several steps. The enveloping flame, due to volatile combustion, acts like a shroud by preventing oxygen from reaching the particle surface for heterogeneous oxidation of char. Following Dryer,[11] the one-step global oxidation of a given species can be written as

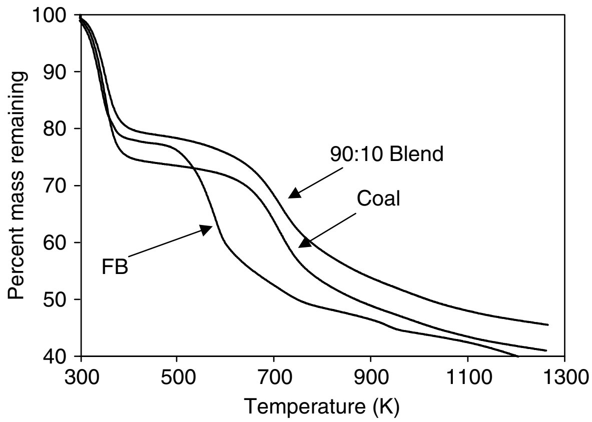


Fig. 6 Thermo-gravimetric analyses of Feedlot Biomass (FB or cattle manure), coal, and 90:10 coal: FB blends.
Source: From Elsevier (see Ref. 4).

Em–Energy Con

$$\text{Fuel} + \upsilon_{O_2} O_2 \xrightarrow{\text{Oxidation}} \upsilon_{CO_2} CO_2 + \upsilon_{H_2O} H_2O \quad (7)$$

$$-\frac{d[\text{Fuel}]}{dt}, \frac{\text{kg}}{\text{m}^3\ \text{sec}} = A \exp\left(\frac{-E}{RT}\right)\left[Y_{\text{fuel}}\right]^a\left[Y_{O_2}\right]^b, \quad (8)$$

where [ ] represents the concentration of species in kg/m^3, Y the mass fraction, A the pre-exponential factor, E the activation energy in kJ/kmole, and a and b the order of reaction; they are tabulated in Bartok and Sarofim for alkanes, ethanol, methanol, benzene and toluene.[11]

Char Reactions

The skeletal char, essentially FC, undergoes heterogeneous reactions with gaseous species. The heterogeneous combustion of carbon or char occurs primarily via one or more of the following reactions:

Reaction I. $C + \frac{1}{2}O_2 \rightarrow CO$

Reaction II. $C + O_2 \rightarrow CO_2$

Reaction III. $C + CO_2 \rightarrow 2\ CO$

Reaction IV. $C + H_2O \rightarrow CO + H_2$

Assuming a first-order reaction for scheme I, the oxygen consumption rate is given as

$$\dot{m}_{O_2} \approx \pi d_p^2 B_I T^n \exp\left(-\frac{E}{R_V T_p}\right)\rho_\infty Y_{O_2,w}. \quad (9)$$

The dominant oxygen transfer mechanism at high temperatures is via reaction I with an E/R (a ratio of activation energy to universal gas constant) of about 26,200 K, where $B_I = 2.3 \times 10^7$ m/s and $n = 0.5$ to 1. Reaction II has an E/R of 20,000 K, and $B_{II} = 1.6 \times 10^5$ m/s. Reaction III, the Boudouard reduction reaction, proceeds with an E/R of about 40,000 K. The reduction reactions, III and IV, may become significant, especially at high temperatures for combustion in boiler burners. Reaction with steam is found to be 50 times faster than CO_2 at temperatures up to 1800°C at 1 bar for 75–100 micron-sized Montana Rosebud char.[12] The combustible gases CO and H_2 undergo gas phase oxidation, producing CO_2 and H_2O.

Ignition and Combustion

Recently, Essenhigh et al. have reviewed the ignition of coal.[13] Volatiles from lignite are known to ignite at $T > 950$ K in fluidized beds. Coal may ignite homogeneously or heterogeneously depending upon size and volatile content.[14,15] A correlation for heterogeneous char ignition temperature is presented by Du and Annamalai, 1994.

Once ignited, the combustion of high volatile coal proceeds in two stages: combustion of VM and combustion of FC. Combustion of VM is similar to the combustion of vapors from an oil drop. The typical total combustion time of 100-micron solid coal particle is on the order of 1 s in boilers and is dominated by the time required for heterogeneous combustion of the residual char particle, while the pyrolysis time ($t_{pyr} = 10^6 (\text{s/m}^2) d_p^2$) is on the order of 1/10th–1/100th of the total burning time. Since bio-solid contains 70%–80% VM (coal contains 10% VM), most of the combustion of volatiles occurs within a short time (about 0.10 s).

For liquid drops and plastics of density ρ_c, simple relations exist for evaluating the combustion rates and times. If the transfer number B is defined as

$$B = \frac{c_p\{T_\infty - T_w\}}{L} + \frac{Y_{O_{2,\infty}}}{\nu_{O_2}}\frac{h_c}{L}, \quad (10)$$

where $T_w \approx$ TBP for liquid fuels; $T_w = T_g$, the temperature of gasification for plastics; L is the latent heat for liquid fuel and $L = q_g$, heat of gasification for plastics; $Y_{O_{2,\infty}}$, is the free-stream oxygen mass fraction; ν_{O_2} is the stoichiometric oxygen mass per unit mass of fuel (typically 3.5 for liquid fuels); and h_c is the lower heating value of fuel; then the burn rate ($\dot{m}$) and time (t_b) for spherical condensates (liquid drops and spherical particles of diameter d_p and density ρ_c) are given by the following expressions:

$$\dot{m} \approx 2\pi \frac{\lambda}{c_p} d_p \ln(1 + B) \quad (11)$$

$$t_b = \frac{d_0^2}{\alpha_c}, \quad (12)$$

where

$$\alpha_c = 8 \frac{\lambda}{c_p} \frac{\ln(1 + B)}{\rho_c} \quad (13)$$

and c_p and λ are the specific heat and thermal conductivity of gas mixture evaluated at a mean temperature (approximately 50% of the adiabatic flame temperature).

The higher the B value, the higher the mass loss rate, and the burn time will be lower. The value of B is about 1–2 for plastics (polymers), 2–3 for alcohols, and 6–8 for pentane to octane. The burn time of plastic waste particles will be about 3–4 times longer than single liquid drops of pentane to octane (≈gasoline) of similar diameter.

COMBUSTION IN PRACTICAL SYSTEMS

The time scales for combustion are on the order of 1000, 10, and 1 ms for coal burnt in boilers and liquid fuels burnt in gas turbines and diesel engines. Coal is burnt on grates in lumped form (larger-sized particles, 2.5 cm or greater with a volumetric intensity on the order of 500 kW/m^3), medium-sized particles in fluidized beds (1 cm or less, 500 kW/m^3), or as suspensions or pulverized fuel (pf; 75 micron or less, 200 kW/m^3) in boilers.

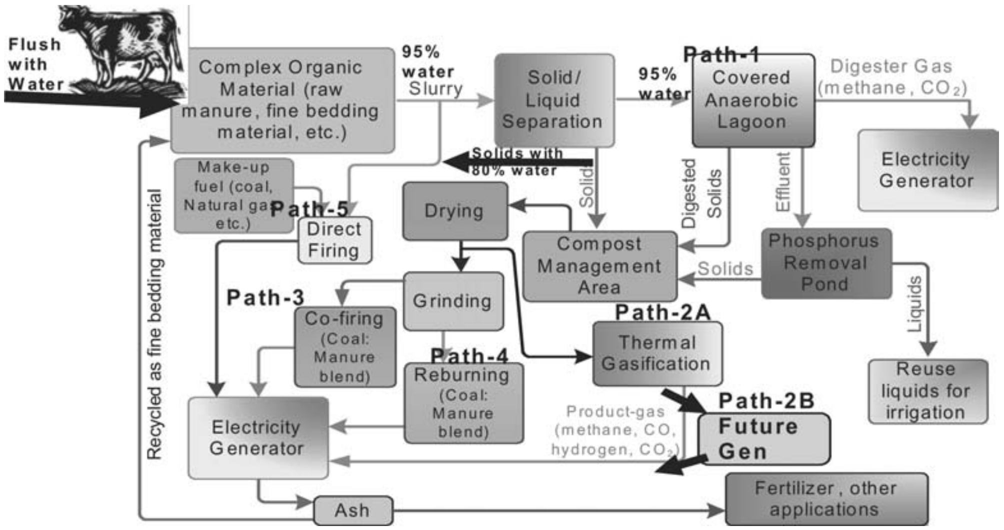


Fig. 7 Flow chart showing several energy conversion options for a typical dairy or cattle operation.

Apart from pyrolysis, gasification, and combustion, another option for energy conversion (particularly if solid fuel is in slurry firm, such as flushed dairy manure), is the anaerobic digestion (in absence of oxygen) to CH_4 (60%) and CO_2 (40%) using psychrophylic (ambient temperature), mesophyllic (95°F) and thermophyllic (135°F) bacteria in digesters.[10] Typical options of energy conversion, indicated in Fig. 7, include anaerobic digestion (path 1, the biological gasification process), thermal gasification with air to produce CO, HC, CO_2 (path 2A) or with steam to produce CO_2 and H_2 (path 2B), cofiring (path 3), reburn (path 4; see "Reburn with Bio-Solids"), and direct combustion (path 5).

Suspension Firing

In suspension-fired boilers, solid fuel is pulverized into smaller particles ($d_p = 75$ μm or less) so that more surface area per unit mass is exposed to the oxidant, resulting in improved ignition and combustion characteristics. Typical boiler burners use swirl burners for atomized oil and pulverized coal firing, while a gas turbine uses a swirl atomizer in highly swirling turbulent flow fields. A swirl burner for pf firing is shown in Fig. 8. The air is divided into a primary air stream which transports the coal (10%–20% of the total air, heated to 70°C–100°C to prevent condensation of vapors and injected at about 20 m/s to prevent settling of the dust, loading dust and gas at a ratio of 1:2) and a secondary air stream (250°C at 60–80 m/s) which is sent through swirl vanes, supplying the remaining oxygen for combustion and imparting a tangential momentum to the air. In wall-fired boilers, burners are stacked above each other on the wall; while in tangential-fired boilers, the burners are mounted at the corners of rectangular furnaces.

Stoker Firing

The uncrushed fuel [fusion temperature <1093°C (2000°F); volatile content >20%; sizes in equal proportions of 19 mm × 12.5 mm (3/4 in. × ½ in.), 6.3 mm × 3.2 mm (½ in. × ¼ in.), 3.2 mm × 3.2 mm (¼ in. × ¼ in.)][16] is fed onto a traveling chain grate below to which primary air is supplied (Fig. 9), which may be preheated to 177°C (350°F) if moisture exceeds 25%. The differential pressure is on the order of 5–8 mm (2–3 in.). The combustible gases are carried into an over-fire region into which secondary air (almost 35% of total air at three levels for low emissions) is fired to complete combustion.

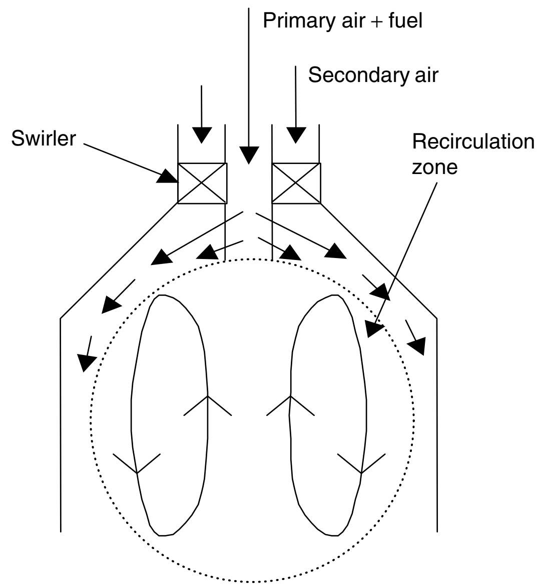


Fig. 8 Pulverized Fuel (pf) fired swirl burner.

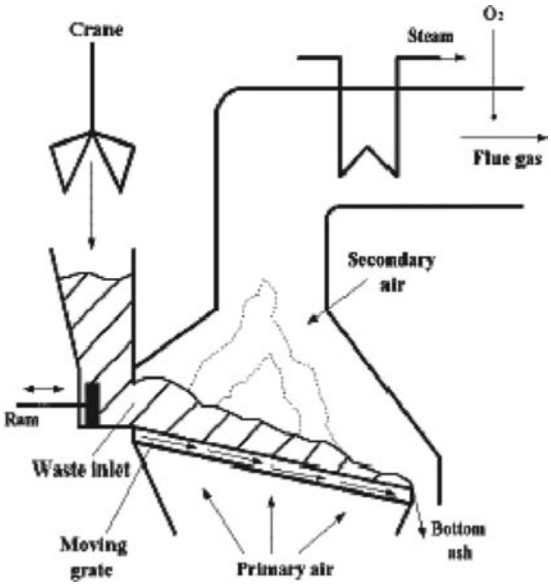


Fig. 9 Schematic of stoker firing.
Source: From Ref. 39.

The over-fire region acts like a perfectly stirred reactor (PSR). It is apparent that solid fuels need not be ground to finer size.

Fixed Bed Combustor

The bed contains uncrushed solid fuels, inert materials (including ash), and processing materials (e.g., limestone to remove SO_2 from gases as sulfates). It is fed with air moving against gravity for complete combustion, but the velocity is low enough that materials are not entrained into the gas streams. Large solid particles can be used.

Fluidized Bed Combustor

When air velocity (V) in fixed bed combustor (FXBC) is increased gradually to a velocity called minimum fluidization velocity V_{mf}, the upward drag force is almost equal to the weight of the particle, so that solids float upward. The bed behaves like a fluid (like liquid water in a tank), i.e., it becomes fluidized. If $V > V_{mf}$, then air escapes as bubbles and is called a bubbling fluidized bed combustor (BFBC). The bed has two phases: the bubble phase, containing gases (mostly oxygen), and the emulsion phase (dense phase, oxygen deficient), containing particles and gas. Many times gas velocity is so high that gaseous combustibles produced within the bed burn above the bed (called free board region), while solids (e.g., char and carbon) burn within the bed. Fluidized Bed Combustor (FBC) is suitable for fuels which are difficult to combust in pf-fired boilers.

Circulating Fluidized Bed Combustor (CFBC)

When air velocity in FBC is increased at velocity $V \gg V_{mf}$, particles are entrained into the gas stream. Since the residence time available to particles for combustion is shorter, unburned particles are captured using cyclones located downstream of the combustor and circulated back to the bed.

The residence time (t_{res}) varies from a low value for pf-fired burners to a long residence time for fixed-bed combustors. The reaction time (t_{reac}) should be shorter than t_{res} so that combustion is complete. The reaction time includes time to heat up to ignition temperature and combustion. The previous section on fuel properties and the homogenous (e.g., CH_4, CO oxidation) and heterogeneous (e.g., carbon oxidation) reaction kinetics can be used to predict t_{reac} or burn time t_b.

COAL AND BIO-SOLIDS COFIRING

General Schemes of Conversion

Most of the previously reviewed combustion systems typically use pure coal, oil, or gas. The same systems require redesign for use with pure biomass fuels. A few of the technologies, which utilize bio-solids as an energy source, are summarized in Annamalai et al.[17] These technologies include direct combustion (fluidized beds), circulating fluidized beds, liquefaction (mostly pyrolysis), onsite gasification for producing low to medium Btu gases, anaerobic digestion (bacterial conversion), and hydrolysis for fermentation to liquid fuels like ethanol.[18,19]

Cofiring

Although some bio-solids have been fired directly in industrial burners as sole-source fuels, limitations arose due to variable moisture and ash contents in bio-solid fuels, causing ignition and combustion problems for direct combustion technologies. To circumvent such problems, these fuels have been fired along with the primary fuels (cofiring) either by directly mixing with coal and firing (2%–15% of heat input basis) or by firing them in between coal-fired burners.[20–24]

Cofiring has the following advantages: improvement of flame stability problems, greater potential for commercialization, low capital costs, flexibility of adaptation of biomass fuels and cost effective power generation, mitigated NO_x emissions from coal-fired boilers, and reduced CO_2 emissions. However, a lower melting point of biomass ash could cause fouling and slagging problems.

Some of the bio-solid fuels used in cofiring with coal are cattle manure,[25,26] sawdust and sewage sludge,[21] switch grass,[20] wood chips,[24,27] straw,[22,28] and refuse-derived fuel (RDF).[21] See Sami et al. for a review of literature on cofiring.[7]

Coal and Agricultural Residues

Sampson et al. reported test burns of three different types of wood chips (20%, HHV from 8320 to 8420 Btu/lb) mixed with coal (10,600 Btu/lb) at a stoker (traveling grate) fired steam plant.[24] The particulate emission in grams per SCF ranged from 0.05 to 0.09. An economic study, conducted for the 125,000 lb/h steam power plant, concluded that energy derived from wood would be competitive with that from coal if more than 30,000 tons of wood chips were produced per year with hauling distances less than 60 mi. Aerts et al. carried out their experiments on cofiring switch grass with coal in a 50-MW, radiant, wall-fired, pulverized coal boiler with a capacity of 180 tons of steam at 85 bar and 510°C (Fig. 10). The NO_x emissions decreased by 20%, since switchgrass contains lesser nitrogen (Table 4).[20] It is the author's hypothesis that a higher VM content of bio-solids results in a porous char, thus accelerating the char combustion process. This is validated by the data from Fahlstedt et al. on the cofiring of wood chips, olive pips and palm nut shells with coal at the ABB Carbon 1 MW Process Test Facility; they found that blend combustion has a slightly higher efficiency than coal-only combustion.[27]

Coal and RDF

Municipal solid waste includes residential, commercial, and industrial wastes which could be used as fuel for production of steam and electric power. MSW is inherently a blended fuel, and its major components are paper (43%); yard waste, including grass clippings (10%); food (10%); glass and ceramics (9%); ferrous materials (6%); and plastics and rubber (5%). Refer to Tables 5 and 6 for analyses. When raw waste is processed to remove non-combustibles like glass and metals, it is called RDF. MSW can decompose in two ways, aerobic and anaerobic. Aerobic decomposition (or composting) occurs when O_2 is present. The composting produces CO_2 and water, but no usable energy products. The anaerobic decomposition occurs in the absence of O_2. It produces landfill gas of 55% CH_4 and 45% CO_2.

Coal and Manure

Frazzitta et al. and Annamalai et al. evaluated the performance of a small-scale pf-fired boiler burner facility (100,000 Btu/h) while using coal and premixed coal-manure blends with 20% manure. Three types of feedlot

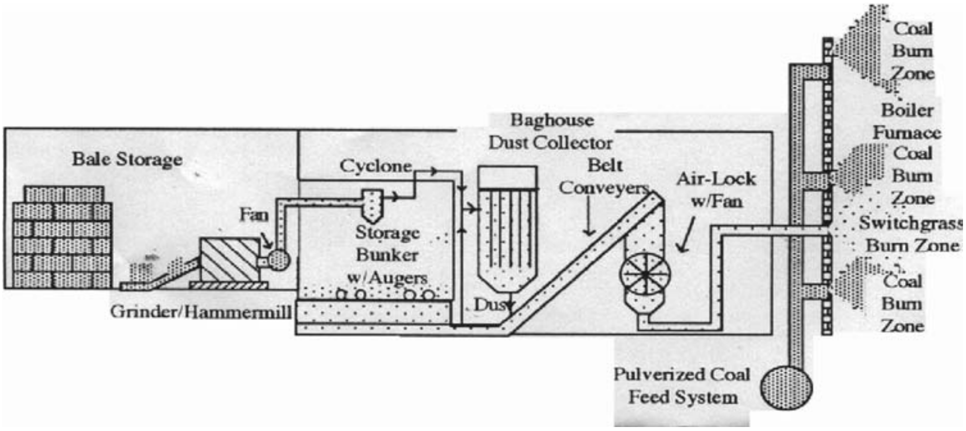


Fig. 10 A cofiring Scheme for coal and biomass (Alternate Fuel Handling Facility at Blount St. Generating Station). Source: From Ref. 20.

Em—Energy Con

manure were used: raw, partially composted, and fully composted. The burnt fraction was recorded to be 97% for both coal and coal-manure blends.[25,26]

NO_x Emissions

During combustion, the nitrogen evolved from fuel undergoes oxidation to NO_x; and this is called fuel NO_x to distinguish it from thermal NO_x, which is produced by oxidation of atmospheric nitrogen. Unlike coal, most of the agricultural biomass being burned is very low in nitrogen content (i.e., wood or crops), but manure has a higher N content than fossil fuels. A less precise correlation exists between cofiring levels on a Btu basis and percent NO reduction under cofiring. The following Eq. (valid between 3 and 22% mass basis cofiring) describes NO_x reduction as a function of cofiring level on a heat input basis:

$$NO_x \text{ Reduction } (\%) = 0.0008\ (COF\%)^2 + 0.0006\ COF\% + 0.0752, \tag{14}$$

where COF% is the percentage of co-firing on a heat input basis. The mechanisms used to reduce NO_x emissions by cofiring vary between cyclone firing and PC firing.

Fig. 11 shows the percentage reduction in NO with percentage cofiring of low-N agricultural biomass fuels. This relationship does not apply to high-N biofuels such as animal manure.

Fouling in Cofiring

Hansen et al. investigated the ash deposition problem in a multi-circulating fluidized bed combustor (MCFBC) fired with fuel blends of coal and wood straw.[25] The Na and K lower the melting point of ash. For ash fusion characteristics see Table 7. Rasmussen and Clausen evaluated the performance of an 80-MW co-generation power plant at Grenaa, Denmark, fired with hard coal and bio-solids (surplus straw from farming). Large amounts of Na and K in straw caused superheater corrosion and combustor fouling.[29] Annamalai et al. evaluated fouling potential when feedlot manure biomass (FB) was cofired with coal under suspension firing.[30] The 90:10 Coal:FB blend resulted in almost twice the ash output compared to coal and ash deposits on heat exchanger tubes that were more difficult to remove than baseline coal ash deposits. The increased fouling behavior with blend is probably due to the higher ash loading and ash composition of FB.

Table 5 Chemical composition of solid waste

	Percent	
Proximate analysis	**Range**	**Typical**
Volatile matter (VM)	30–60	50
Fixed carbon (FC)	5–10	8
Moisture	10–45	25
Ash	10–30	25

	Percent by mass (dry basis)					
Ultimate analysis	**C**	**H**	**O**	**N**	**S**	**Ash**
Yard wastes	48	6	38	3	0.3	4.7
Wood	50	6	43	0.2	0.1	0.7
Food wastes	50	6	38	3	0.4	2.6
Paper	44	6	44	0.3	0.2	5.5
Cardboard	44	6	44	0.3	0.2	5.5
Plastics	60	7	23			10
Textiles	56	7	30	5	0.2	1.8
Rubber	76	10		2		12
Leather	60	9	12	10	0.4	8.6
Misc. organics	49	6	38	2	0.3	4.7
Dirt, ashes, etc.	25	3	1	0.5	0.2	70.3

Table 6 Heat of combustions of municipal solid waste components

	Inerts (%)		Heating values (kJ/kg)	
Component	**Range**	**Typical**	**Range**	**Typical**
Yard wastes	2–5	4	2,000–19,000	7,000
Wood	0.5–2	2	17,000–20,000	19,000
Food wastes	1–7	6	3,000–6,000	5,000
Paper	3–8	6	12,000–19,000	17,000
Cardboard	3–8	6	12,000–19,000	17,000
Plastics	5–20	10	30,000–37,000	33,000
Textiles	2–4	3	15,000–19,000	17,000
Rubber	5–20	10	20,000–28,000	23,000
Leather	8–20	10	15,000–20,000	17,000
Misc. organics	2–8	6	11,000–26,000	18,000
Glass	96–99	98	100–250	150
Tin cans	96–99	98	250–1,200	700
Nonferrous	90–99	96		
Ferrous metals	94–99	98	250–1,200	700
Dirt, ashes, etc.	60–80	70	2,000–11,600	7,000

GASIFICATION OF COAL AND BIO-SOLIDS

Gasification is a thermo-chemical process in which a solid fuel is converted into a gaseous fuel (primarily consisting of HC, H_2 and CO_2) with air or pure oxygen used for partial oxidation of FCs. The main products during gasification are CO and H_2, with some CO_2, N_2, CH_4, H_2O, char particles, and tar (heavy hydrocarbons). The oxidizers used for the gasification processes are oxygen, steam, or air. However, for air, the gasification yields a low-Btu gas, primarily caused by nitrogen dilution present in the supply air. Syngas ($CO+H_2$) is produced by reaction of biomass with steam. The combustible product, gas, can be used as fuel burned directly or with a gas turbine to produce electricity; or used to make chemical feedstock (petroleum refineries). However, gas needs to be cleaned to remove tar, NH_3, and sulfur compounds. The integrated gasification combined cycle (IGCC) (Fig. 12),

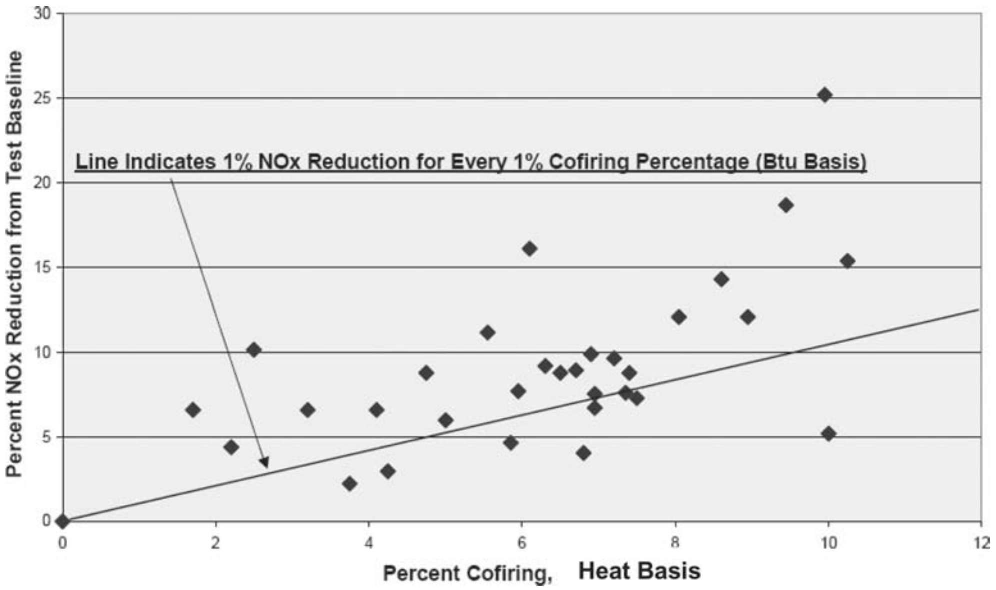


Fig. 11 NO_x reduction due to cofiring with low-N agricultural residues.
Source: From Technical Advisory Committee (see Ref. 40).

Em—Energy Con

for combined heat and power (CHP), and traditional boilers use combustible gases from gasifiers for generation of electric power.

Typically in combined cycles, gaseous or liquid fuel is burnt in gas turbine combustors. High-temperature products are expanded in a gas turbine for producing electrical power; a low-temperature (but still hot) exhaust is then used as heat input in a boiler to produce low-temperature steam, which then drives a steam turbine for electrical power. Therefore, one may use gas as a topping cycle medium, while steam is used as fluid for the bottoming cycle. The efficiency of a combined cycle is on the order of 60%, while a conventional gas turbine cycle has an efficiency of 42%.[31] Commercial operations include a 250-MW IGCC plant at Tampa, Florida, operating since 1996; a 340-MW plant at Negishi, Japan, since 2003; and a 1200-MW GE-Bechtel plant under construction in Ohio for American Electric Power, to start in 2010.[31]

There are three basic gasification reactor types: (i) fixed-bed gasifiers (Fig. 13); (ii) fluidized-bed gasifiers, including circulating-bed (CFB) or bubbling-bed; and (iii) entrained-flow gasifiers. The principles of operation are similar to those of combustors except that the air supplied is much below stoichiometric amounts, and instead of a combination of steam, air and CO_2, air can also be used. The oxidant source could also include gases other than air, such as air combined with steam in Blasiak et al.[32]

Table 7 Ash fusion behavior and ash composition, fusion data: ASTM D-188

	FB	PRB Coal	Blend
Ash Fusion, (reducing)			
Initial deformation, IT, °C (°F)	1140 (2090)	1130 (2060)	NA
Softening, °C (°F)	1190 (2170)	1150 (2110)	NA
Hemispherical, HT, °C (°F)	1210 (2210)	1170 (2130)	NA
Fluid, °C (°F)	1230 (2240)	1200 (2190)	NA
Ash fusion, (oxidizing)			
Initial deformation, IT, °C (°F)	1170 (2130)	1190 (2180)	NA
Softening, °C (°F)	1190 (2180)	1200 (2190)	NA
Hemispherical, HT, °C (°F)	1220 (2230)	1210 (2210)	NA
Fluid, °C (°F)	1240 (2270)	1280 (2330)	NA
Slagging Index, Rs, °C (°F)	1160 (2120)	1140 (2090)	
Slagging classification	High	Severe	
Ash composition (wt%)			
SiO_2	53.63	36.45	43.56
Al_2O_3	5.08	18.36	12.87
Fe_2O_3	1.86	6.43	4.54
TiO_2	0.29	1.29	0.88
CaO^+	14.60	19.37	17.40
MgO^+	3.05	3.63	3.39
Na_2O^+	3.84	1.37	2.39
K_2O^+	7.76	0.63	3.58
P_2O_5	4.94	0.98	2.62
SO_3	3.71	10.50	7.69
MnO_2	0.09	0.09	0.09
Sum	98.84	99.11	99.00
Volatile Oxides	30.77	28.25	
Basic oxides	32.73	35.51	
Silica ratio	0.73	0.53	
Na_2O+K_2O	11.60	2.00	5.97
Inherent Ca/S Ratio	6.71	1.86	2.48
kg alkali (Na_2O+K_2O)/GJ	5.37	0.06	0.29

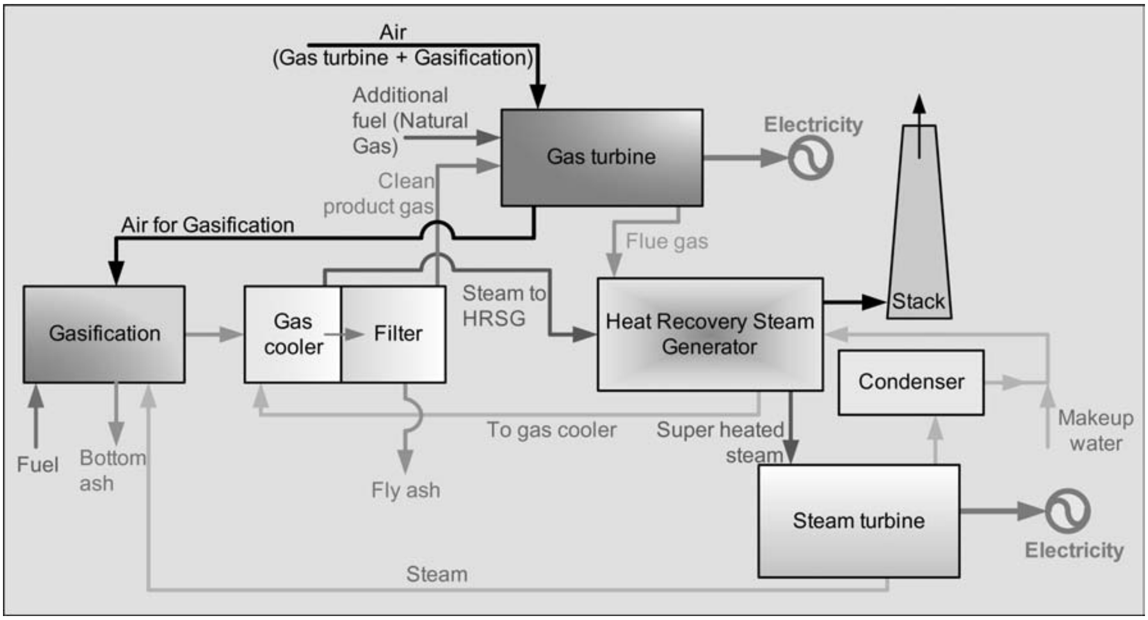


Fig. 12 Fluidized-bed gasification for Integrated Gasification Combined Cycle (IGCC) Process.

FUTUREGEN

FutureGen is a new U.S. initiative to build the world's first integrated CO_2 sequestration and H_2 production research power plant using coal as fuel. The technology shown in Fig. 14 employs modern coal gasification technology using pure oxygen, resulting in CO, C_nH_m (a hydrocarbon), H_2, HCN, NH_3, N_2, H_2S, SO_2, and other combinations which are further reacted with steam (reforming reactions) to produce CO_2 and H_2. The bed materials capture most of the harmful N and S compounds followed by gas-cleaning systems; the CO_2 is then sequestered and H_2 is used as fuel, using either combined cycle or fuel cells for electricity generation or sold as clean transportation fuel. With partial oxidation of gasification products and char supplying heat for pyrolysis and other endothermic reactions (i.e, net zero external heat supply in gasifier), the overall gasification reaction can be represented as follows for 100 kg of DAF Wyoming coal:

Reaction V $$C_{6.3}H_{4.4}N_{0.076}O_{1.15}S_{0.014} + 8.14\ H_2O(\ell) + 1.67\ O_2 \rightarrow 6.3\ CO_2 + 10.34\ H_2 + 0.038\ N_2 + 0.014\ SO_2$$

It is apparent that the FutureGen process results in enhanced production of H_2, using coal as an energy source to strip H_2 from water. For C–H–O fuels, it can be shown that theoretical H_2 production (N_{H_2}) in moles for an empirical fuel CH_hO_o is given as $\{0.4115\,h - 0.6204\,o + 1.4776\}$ under the above conditions. For example, if glucose $C_6H_{12}O_6$ is the fuel, then empirical formulae is CH_2O; thus, with $h=2$, $o=1$, $N_{H_2}=1.68$ kmol, using atom balance, $CH_2O + 0.68\ H_2O(\ell) + 0.16\ O_2 \rightarrow CO_2 + 1.68\ H_2$.

REBURN WITH BIO-SOLIDS

NO_x is produced when fuel is burned with air. The N in NO_x can come both from the nitrogen-containing fuel

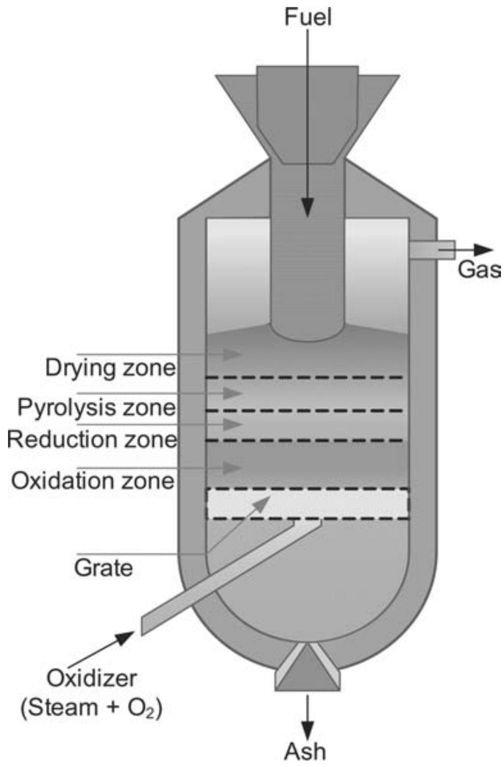


Fig. 13 Updraft fixed-bed gasifier.

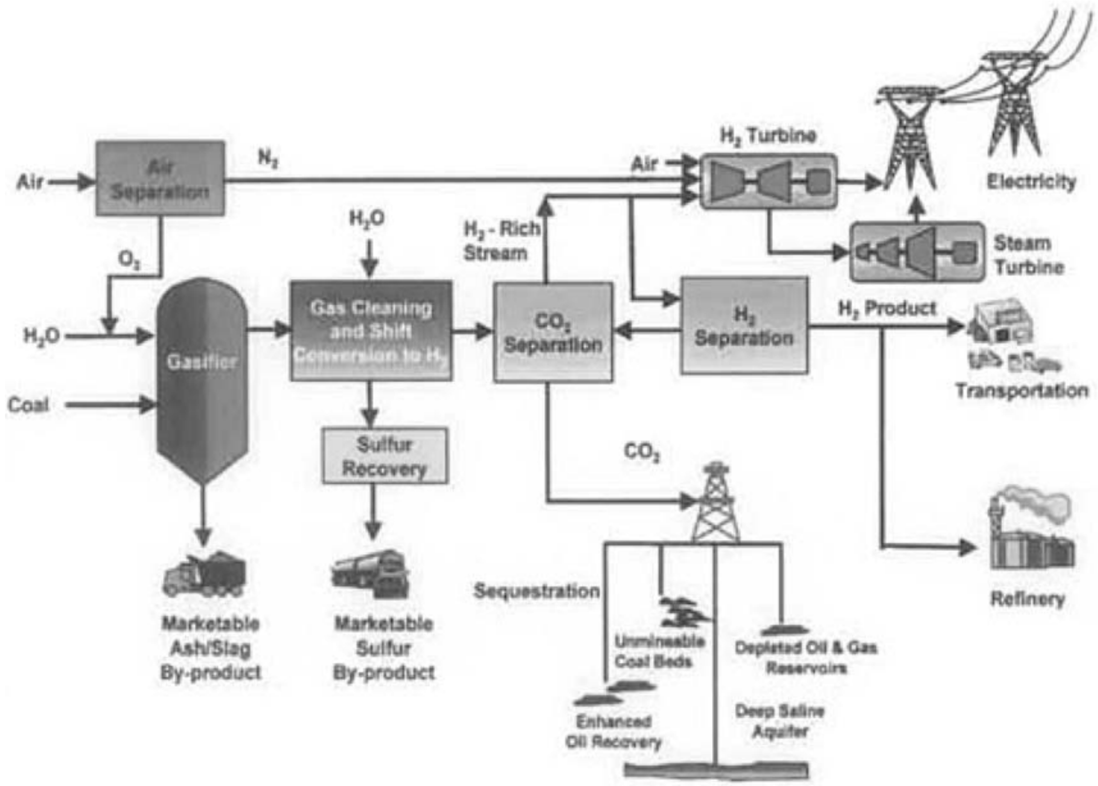


Fig. 14 FutureGen layout.
Source: From http://www.fe.doe.gov.

compounds (e.g., coal, biomass, plant residue, animal waste) and from the N in the air. The NO_x generated from fuel N is called fuel NO_x, and NO_x formed from the air is called thermal NO_x. Typically, 75% of NO_x in boiler burners is from fuel N. It is mandated that NO_x, a precursor of smog, be reduced to 0.40–0.46 lb/mmBtu for wall and tangentially fired units under the Clean Air Act Amendments (CAAA). The current technologies developed for reducing NO_x include combustion controls (e.g., staged combustion or low NO_x burners (LNB), reburn) and post-combustion controls (e.g., Selective Non-Catalytic Reduction, SNCR using urea).

In reburning, additional fuel (typically natural gas) is injected downstream from the primary combustion zone to create a fuel rich zone (optimum reburn stoichiometric ratio (SR), usually between SR 0.7 and 0.9), where NO_x is reduced up to 60% through reactions with hydrocarbons when reburn heat input with CH_4 is about 10%–20%. Downstream of the reburn zone, additional air is injected in the burnout zone to complete the combustion process. A diagram of the entire process with the different combustion zones is shown in Fig. 15. There have been numerous studies on reburn technology found in literature, with experiments conducted, and the important results summarized elsewhere.[33] Table 8 shows the percentages of reduction and emission obtained with coal or gas reburn in coal-fired installations and demonstration units.

The low cost of biomass and its availability make it an ideal source of pyrolysis gas, which is a more effective reburn fuel than the main source fuel, which is

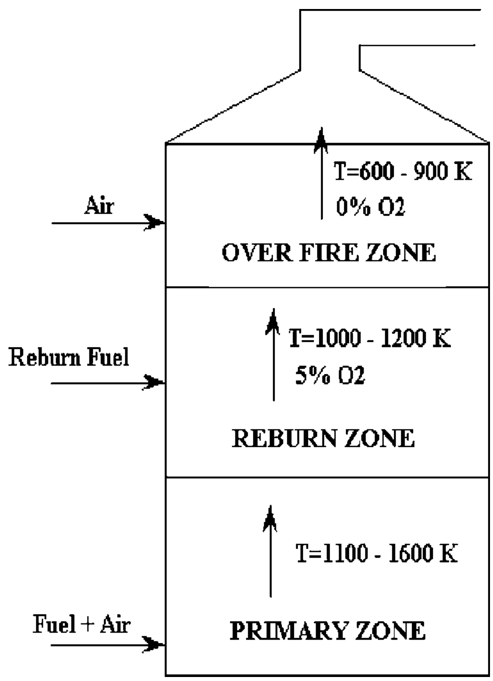


Fig. 15 Schematic of reburn process.

Em–Energy Con

Table 8 Percentage reduction in NO_x: demonstration and/or operating reburn installations on coal-fired boilers in the United States

Type of Burner	% Reburn Heat in	% Reduction	NOx with Reburn lb/mmBtu[a]
Gas reburning			
Tangential	18	50–67	0.25
Cyclone	20–23	58–60	0.39–0.56
Wall without LNB	18	63	0.27
Coal reburn			
Cyclone (micronized)	30 (17)	52 (57)	0.39 (0.59)
Tangential(micron)with LNB	14	28	0.25

LNB: Low NO_x Burners.
[a] 1 lb per mmBtu = 0.430 kg/GJ.
Source: From U.S. Department of Energy (see Ref. 38).

typically coal. Recently, animal manure has been tested as a reburn fuel in laboratory scale experiments. A reduction of a maximum of 80% was achieved for pure biomass, while the coal experienced a reduction of between 10 and 40%, depending on the equivalence ratio.[34] It is believed that the greater effectiveness of the feedlot biomass is due to its greater volatile content on a DAF basis and its release of fuel nitrogen in the form of NH_3, instead of HCN.[35]

ACKNOWLEDGMENTS

Most of this work was supported in part by the U.S. Department of Energy of Pittsburgh, PA, the DOE of Golden, CO, the USDA, the Texas Advanced Technology Program, and the Texas Commission on Environmental Quality (TCEQ) of Austin, TX, through the Texas Engineering Experiment Station (TEES), Texas A&M University, and College Station, TX.

REFERENCES

1. U.S. DOE, Energy Information Administration (www.eia.doe.gov/, accessed April 12, 2005).
2. Priyadarsan, S.; Annamalai, K.; Sweeten, J.M.; Holtzapple, M.T.; Mukhtar, S. Co-gasification of blended coal with feedlot and chicken litter biomass, Proceedings of the 30th Symposium (International) on Combustion; The Combustion Institute: Pittsburgh, PA, 2005; Vol. 30, 2973–2980.
3. Ebeling, J.M.; Jenkins, B.M. Physical and chemical properties of biomass fuels. Trans. ASAE **1985**, *28* (3), 898–902.
4. Sweeten, J.M.; Annamalai, K.; Thien, B.; McDonald, L. Co-firing of coal and cattle feedlot biomass (FB) fuels, part I: feedlot biomass (cattle manure) fuel quality and characteristics. Fuel **2003**, *82* (10), 1167–1182.
5. Boie, W. Wiss z Tech. Hochsch. Dresden **1952/1953**, *2*, 687.
6. Annamalai, K.; Sweeten, J.M.; Ramalingam, S.C. Estimation of the gross HVs of biomass fuels. Trans. Soc. Agric. Eng. **1987**, *30*, 1205–1208.
7. Sami, M.; Annamalai, K.; Wooldridge, M. Co-firing of coal and biomass fuel blends. Prog. Energy Combust. Sci. **2001**, *27*, 171–214.
8. Annamalai, K.; Puri, I.K. 2006, *Combustion Science and Engineering*, Taylor and Francis: Orlando, FL, 2006.
9. Tillman, D.A.; Rossi, A.J.; Kitto, W.D. *Wood Combustion: Principles: Processes and Economics*; Academic press: New York, 1981.
10. Sweeten, J.M.; Heflin, K.; Annamalai, K.; Auvermann, B.; Collum, Mc; Parker, D.B. *Combustion-Fuel Properties of Manure Compost from Paved vs Un-paved Cattle feedlots*; ASABE 06-4143: Portland, OR, July 9–12, 2006.
11. Bartok, W., Sarofim, A.F. 1991, Fossil Fuel Combustion, chapter 3: FL Dryer, John Wiley, and Hobokan, NJ. pp. 121–214.
12. Howard, J.B.; Sarofim, A.F. Gasification of coal char with CO_2 and steam at 1200–1800°C. *Energy Lab Report*; Chemical Engineering; MIT: MA, 1978.
13. Essenhigh, R.H.; Misra, M.K.; Shaw, D.W. Ignition of coal particles: a review. Combust. Flame **1989**, *77*, 3–30.
14. Annamalai, K.; Durbetaki, P. A theory on transition of ignition phase of coal particles. Combust. Flame **1977**, *29*, 193–208.
15. Du, X.; Annamalai, K. The transient ignition of isolated coal particles. Combust. Flame **1994**, *97*, 339–354.
16. Johnson, N. Fundamentals of Stoker Fired Boiler Design and Operation, CIBO Emission Controls Technology Conference, July 15–17, 2002.
17. Annamalai, K.; Ibrahim, Y.M.; Sweeten, J.M. Experimental studies on combustion of cattle manure in a fluidized bed combustor. Trans. ASME, J. Energy Res. Technol. **1987**, *109*, 49–57.
18. Walawender, W.P.; Fan, L.T.; Engler, C.R.; Erickson, L.E. Feedlot manure and other agricultural wastes as future material and energy resources: II. Process descriptions. *Contrib.30*, Deptartment of Chemical

Em–Energy Con

Engineering, Kansas Agricultural Experiment Station: Manhattan, KS, 30, 1973.

19. Raman, K.P.; Walawander, W.P.; Fan, L.T. Gasification of feedlot manure in a fluidized bed: effect of temperature. Ind. Eng. Chem. Proc. Des. Dev. **1980**, *10*, 623–629.
20. Aerts, D.J.; Bryden, K.M.; Hoerning, J.M.; Ragland, K.W. *Co-firing Switchgrass in a 50 MW Pulverized Coal Boiler*, Proceedings of the 59th Annual American Power Conference, Chicago, IL, 1997; Vol. 50(2), 1180–1185.
21. Abbas, T.; Costen, P.; Kandamby, N.H.; Lockwood, F.C.; Ou, J.J. The influence of burner injection mode on pulverized coal and biosolid co-fired flames. Combust. Flame **1994**, *99*, 617–625.
22. Siegel, V.; Schweitzer, B.; Spliethoff, H.; Hein, K.R.G. Preparation and co-combustion of cereals with hard coal in a 500 kW pulverized-fuel test unit. Biomass for energy and the environment, *Proceedings of the 9th European Bioenergy Conference*, Copenhagen, DK, 24–27 June, 1996; 2, 1027–1032.
23. Hansen, L.A.; Michelsen, H.P.; Dam-Johansen, K. Alkali metals in a coal and biosolid fired CFBC—measurements and thermodynamic modeling, *Proceedings of the 13th International Conference on Fluidized Bed Combustion*, Orlando, FL, May 7–10, 1995; 1, 39–48.
24. Sampson, G.R.; Richmond, A.P.; Brewster, G.A.; Gasbarro, A.F. Co-firing of wood chips with coal in interior Alaska. Forest Prod. J. **1991**, *41* (5), 53–56.
25. Frazzitta, S.; Annamalai, K.; Sweeten, J. Performance of a burner with coal and coal: feedlot manure blends. J. Propulsion Power **1999**, *15* (2), 181–186.
26. Annamalai, K.; Thien, B.; Sweeten, J.M. Co-firing of coal and cattle feedlot biomass (FB) fuels part II: performance results from 100,000 Btu/h laboratory scale boiler burner. Fuel **2003**, *82* (10), 1183–1193.
27. Fahlstdedt, I.; Lindman, E.; Lindberg, T.; Anderson, J. Co-firing of biomass and coal in a pressurized fluidized bed combined cycle. Results of pilot plant studies, *Proceedings of the 14th International conference on Fluidized Bed Combustion*, Vancouver, Canada, May 11–14, 1997; 1, 295–299.
28. Van Doorn, J., Bruyn, P., Vermeij, P. Combined combustion of biomass, fluidized sewage sludge and coal in an atmospheric Fluidized bed installation, Biomass for energy and the environment, *Proceedings of the 9th European Bioenergy Conference*, Copenhagen, DK, June 24–27, 2, 1996; 1007–1012.
29. Rasmussen, I.; Clausen, J.C. ELSAM strategy of firing biosolid in CFB power plants, *Proceedings of the 13th International Conference on Fluidized Bed Combustion*, Orlando, FL, May 7–10, 1, 1995; 557–563.
30. Annamalai, K.; Sweeten, J.; Freeman, M.; Mathur, M.; O'Dowd, W.; Walbert, G.; Jones, S. Co-firing of coal and cattle feedlot biomass (FB) fuels, part III: fouling results from a 500,000 Btu/hr pilot plant scale boiler burner. Fuel **v.82,10**, *82* (10), 1195–1200, 2003.
31. Langston, L.S. 2005 New Horizons. Power and Energy vol 2, 2, June 2005.
32. Blasiak, W.; Szewczyk, D.; Lucas, C.; Mochida, S. *Gasification of Biomass Wastes with High Temperature Air and Steam*, Twenty-First International Conference on Incineration and Thermal Treatment Technologies, New Orleans, LA, May 13–17, 2002.
33. Thien, B.; Annamalai, K. National Combustion Conference, Oakland, CA, March 25–27, 2001.
34. Arumugam, S.; Annamalai, K.; Thien, B.; Sweeten, J. Feedlot biomass co-firing: a renewable energy alternative for coal-fired utilities, Int. Natl J. Green Energy, vol. 2, No. 4, 409–419, 2005.
35. Zhou, J.; Masutani, S.; Ishimura, D.; Turn, S.; Kinoshita, C. Release of fuel-bound nitrogen during biomass gasification. Ind. Eng. Chem. Res. **2000**, *39*, 626–634.
36. Annamalai, K.; Puri, I.K. *Advanced Thermodynamics*; CRC Press: Boca Raton, FL, 2001.
37. Priyadarsan, S.; Annamalai, K.; Sweeten, J.M.; Holtzapple, M.T.; Mukhtar, S. Waste to energy: fixed bed gasification of feedlot and chicken litter biomass. Trans. ASAE **2004**, *47* (5), 1689–1696.
38. DOE, *Reburning technologies for the control of nitrogen oxides emissions from coal-fired boilers* Topical Report, No 14, U.S. Department of Energy, May, 1999.
39. Loo, S.V.; Kessel, R., 2003, Optimization of biomass fired grate stroker systems, EPRI/IEA Bioenergy Workshop, Salt Lake City, Utah, February 19, 2003.
40. Grabowski, P. *Biomass Cofiring, Office of the Biomass Program*, Technical Advisory Committee, Washington, DC, March 11, 2004.

Energy Efficiency: Developing Countries

U. Atikol
H.M. Güven
Energy Research Center, Eastern Mediterranean University, Magusa, Northern Cyprus

Abstract
Statistics and projections show that not only the rate of energy consumption, but also the carbon emissions of developing countries are rising very fast. The rate of increase in the capacity needs in developing countries can be decreased in two steps. First, the component associated with the old infrastructure should be dealt with; and second, the application of end-use energy efficiency measures should be put in effect. The energy efficiency technologies that are available in the industrialized countries may not always be feasible for transfer to developing countries. Readily available economic and social indicators, supported by more detailed end-use research, can be used to determine which technologies are suitable for a given developing country. Case studies performed in Northern Cyprus and Turkey show that in Northern Cyprus, transfer of these technologies would be more successful in the residential and commercial sectors, whereas in Turkey, they would be more feasible for the industrial sector.

INTRODUCTION

The quest for energy efficiency dates back to 1973, at which time the Oil Producing and Exporting Countries (OPEC) placed an embargo on crude oil during the war in the Middle East. Until that year, most developed countries had experienced decades of low energy prices and plentiful fuel supplies; consequently, high and growing per capita use of energy was of little concern to most governments. In 1979, there was another interruption in supplies when U.S. President Jimmy Carter announced a ban on oil imports from Iran after the hostage crisis began at the U.S. embassy in Tehran. Subsequently, rapidly rising energy costs and interruptions in supplies forced re-evaluation of existing policies. In most developed countries, conservation and efficiency improvements to energy systems became an important component of energy policy, due not only to the realization that the world is vulnerable to interference with its energy supplies, but also to awareness that excessive use of fossil fuels is causing vital damage to the world ecosystem.

In the United States before the 1973 crises, both primary energy and electricity consumption increased at almost the same rate as the Gross National Product (GNP). (The Gross National Product is the dollar value of all goods and services produced in a nation's economy, including goods and services produced abroad). After that year, however, high oil prices led to progressive energy policies, and the demand was halted (see Fig. 1). Between 1973 and 1986, the rate of growth of electricity use decreased, growing only 2.5% per year, or 3.2% per year less than projected by pre1973 trends.[1] In 1986, projected electricity use was 50% higher than actual electricity use, indicating a savings of 1160 TWh. The gap between the two trends is referred to as JAWS, coined by Rosenfeld.

Since 1970, the growth of electricity consumption has been around 3.5% for the countries in the Organization for Economic Cooperation and Development (OECD), (Organization for Economic Cooperation and Development countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, The Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.) while that of developing countries has been 8.2%. For the OECD countries, this growth is almost parallel with the growth of their Gross Domestic Product (GDP). (Gross Domestic Product is a measure of the value of all goods and services produced by the economy. Unlike GNPs, GDP includes only the values of goods and services earned by a nation within its boundaries). In the developing countries, however, electricity generation rose much faster than the GDP. While the total GDP (in U.S. dollars) increased slightly over twofold between 1970 and 1989, electricity generation grew more than fourfold.[2]

It was rather fortunate that there was ample scope for improvements in energy efficiency. Insulation and other measures, such as using more efficient air conditioners, were applied to reduce space heating and air conditioning requirements. The wide use of household appliances meant a large market for more energy-efficient electric motors, compressors, water heaters, and other equipment. In transportation, improvement in fuel use was also achieved for new vehicles. In industrialized countries,

Keywords: Developing countries; Energy efficiency; Demand-side management; Energy planning; Technology transfer.

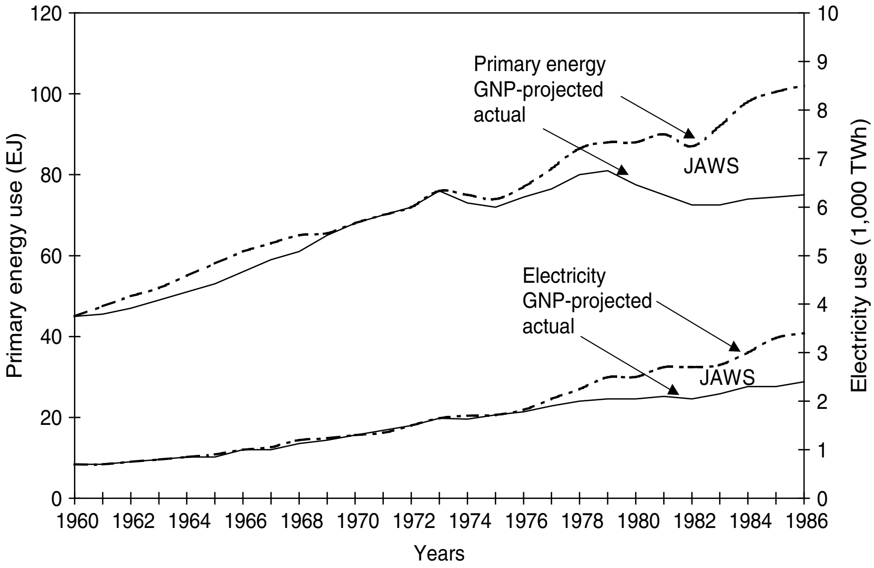


Fig. 1 U.S. total primary energy and electricity use immediately after 1973: actual and GNP projected. The graph shows the effect of energy efficiency applications between 1973 and 1986.
Source: From Island Press (see Ref. 1).

several mechanisms were developed to implement the changes; two important ones were the introduction of utility demand-side management (DSM) and the establishment of energy service companies (ESCOs) backed up with accurate data collection and improved codes and standards. Recently, distributed power and renewable energy (RE) technologies have been promoted by the introduction of new legislation.

In developing countries, these mechanisms do not exist, mainly due to lack of institutional formation, technical expertise, and sufficient infrastructure. In this article, there will be an attempt to evaluate these problems with suggestions for possible solutions.

TRENDS IN WORLDWIDE ENERGY DEMAND

In a report prepared by the Energy Information Administration,[3] the International Energy Outlook 2000 (IEO2000), much of the growth in worldwide energy use is projected for the developing world (Fig. 2a). In particular, energy demand in developing Asia and Central and South America is projected to more than double between 1997 and 2020. Both regions are expected to sustain energy demand growth of more than 3% annually throughout the forecast, accounting for more than 50% of the total projected increment in world energy consumption and 83% of the increment for the developing world alone. World carbon emissions are projected to rise from 6.2 billion metric tons in 1997 to 8.1 billion metric tons in 2010 and to 10 billion metric tons in 2020, according to the IEO2000. This analysis does not take into account the potential impact of the Kyoto Protocol. (The Kyoto Climate Change Protocol is a treaty signed by 83 countries and the European Union that requires reductions or limits to the growth of carbon emissions within the Annex I countries between 2008 and 2012. The Annex I countries under the protocol are Australia, Bulgaria, Canada, Croatia, the European Union, Iceland, Japan, Liechtenstein, Monaco, New Zealand, Norway, Romania, Russia, Switzerland, Ukraine, and the United States. Turkey and Belarus are Annex I countries that did not commit to quantifiable emissions targets under the protocol). In this forecast, world carbon emissions will exceed their 1990 levels by 40% in 2010 and by 72% in 2020. Emissions in the developing countries accounted for about 28% of the world total in 1990, but they are projected to make up 44% of the total by 2010 and nearly 50% by 2020. As a result, even if the Annex I countries were able to meet the emissions limits or reductions prescribed in the Kyoto Protocol, worldwide carbon emissions still would grow substantially. The increase is expected to be caused both by rapid economic expansion, accompanied by growing demand for energy, and by continued heavy reliance on coal (the most carbon intensive of the fossil fuels), particularly in developing Asia.

There is an expected increase in electricity consumption worldwide by 76% in the IEO2000 reference case, from 12 trillion kWh in 1997 to 22 trillion kWh in 2020. Long-term growth in electricity consumption is expected to be strongest in the developing countries of Asia, followed by those of Central and South America. Those two regions alone account for 52% of the world's net electricity consumption increment in the IEO2000 reference case (Fig. 2b). Rapid growth in population and income, along with greater industrialization and more

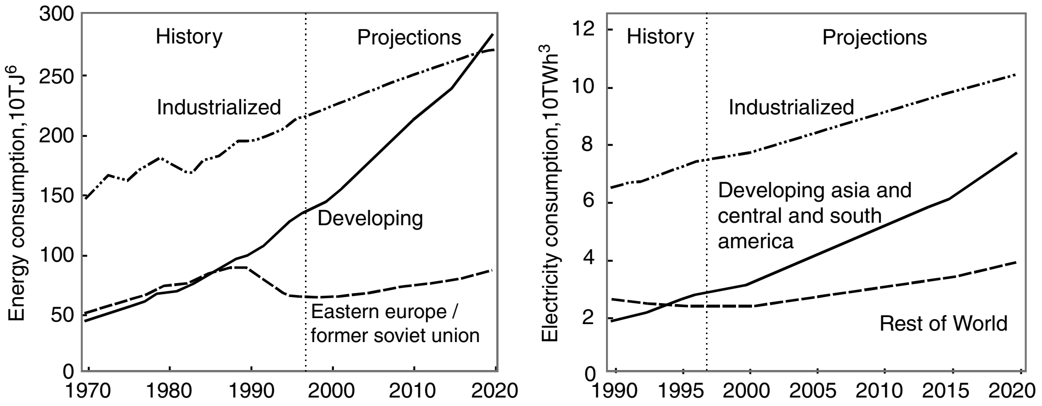


Fig. 2 World energy and net electricity consumption by region, 1970–2020.
Source: From U.S. Department of Energy (see Ref. 3).

household electrification, are responsible for the increase.

Developing countries need energy to raise productivity and improve the living standards of their populations. Traditionally, developing countries have addressed their energy needs by expanding their supply base, with little attention to the efficient use of energy. This approach has been raising serious financial, institutional, and environmental problems. The magnitude of these problems has been underlining the need for improving the efficiency with which energy is currently used and produced in developing countries.

Factors contributing to the rapidly rising energy consumption in developing countries include population growth, economic growth, and increased consumer demand.

Population Growth in Developing Countries

The world population is just over 6 billion and is projected by the U.S. Bureau of Census[4] to reach 9.1 billion in 2050. Over the next two decades, the population of the developing world is projected to increase by nearly 1 billion, to almost 6 billion total, while that of the industrialized countries will increase by only 50 million, to 1 billion total.[3] By 2050, the population of the less-developed regions could reach 7.75 billion or more.[5] Developing countries then could account for 85% of the global population. The increase in population between 1998 and 2050 for the developing countries would be 64%, and this alone would account for a large increase in their energy consumption by 2050, even if per capita consumption remained at current levels.

Economic Growth in Developing Countries

Securing higher living standards for this rising population requires rapid economic growth, further increasing the demand for energy services. The steady increase in per capita energy consumption in developing countries is due merely to economic growth, which includes urbanization, increased use of commercial fuels instead of traditional biomass fuels (such as wood, crop residues, and animal dung), and increased use of energy-intensive materials. When annual per capita energy consumptions of the industrialized and developing countries are compared, it is observed that the gap between them is more than 4.5 million tons of oil equivalent (TOE).[3] This shows that the developing countries have a long way to go to increase their living standards to those of industrialized countries. Closing this gap will lead to astronomic increases in the world energy consumption. This is the inevitable and unavoidable eventuality that may lead to catastrophic environmental consequences. Therefore, utilizing energy as efficiently as possible and trying to use cleaner sources of energy is absolutely essential.

Accelerated Consumer Demand

In recent years, modern manufacturing techniques and improved materials have sharply lowered the real costs of consumer goods such as radios, refrigerators, and televisions. In the 1990s, for example, a study[6] revealed that the real cost of refrigerators decreased by a factor of 5 between 1950 and 1990. Similarly, in Northern Cyprus, due to the decreasing cost of air conditioners by a factor of 2.5 between 1996 and 2006, not only the number of residences, but also the number of small to medium-size enterprises using them increased dramatically. Global distribution systems have also increased the accessibility of these appliances. Thus, people in developing countries can purchase these goods at a far earlier stage in the development cycle (as measured by per capita GDP) than did people in today's industrial countries. The rapidly increasing use of these consumer goods has a strong impact on residential electricity use, creating additional demand at peak times.

ENERGY AND DEVELOPING COUNTRIES

Developing countries need energy to raise productivity and improve the living standards of their populations. Traditionally, developing countries have addressed their energy needs by expanding their supply base, with little attention to the efficient use of energy, which led to the need to expand the supply base frequently. This approach, however, raises serious financial, institutional, and environmental problems. The magnitude of these problems underlines the need for improving the efficiency with which energy is currently used and produced in developing countries.

The magnitude of funding needed for the required power supply projects is growing so fast that these projects have little chance of being mobilized. The power sectors in developing countries frequently experience a wide range of institutional problems, including excessive staffing, inadequate management, weak planning, poor maintenance, deficient financial monitoring, and few incentives to improve efficiency of operation.[7] This raises questions about the ability of this key sector to continue expanding rapidly even if financial resources were available. The continuous increase in energy consumption (fossil-fuel combustion in modern industry, transport, and electricity generation) causes air pollution and contaminates water supplies and land. On the other hand, nonfossil energy sources may also cause environmental damage. Hydroelectric development often involves the flooding of vast areas of land, with a resulting loss of agricultural land, human settlements, fish production, forests, wildlife habitats, and species diversity. Therefore, using energy as efficiently as possible should be an essential part of the planning process.

ENERGY EFFICIENCY IMPLEMENTATION BARRIERS

As the demand for electricity increased and the global climate change issues caused governmental concern, energy efficiency attracted attention in some developing countries. Difficulties were experienced in the implementation of these energy efficiency measures and programs, however. The implementation barriers were identified as follows:[8,9]

- Lack of central coordination and institutional formation
- Lack of awareness and general misinformation
- Lack of technical information and expertise
- Inappropriate pricing policies
- Lack of capital
- Lack of appropriate laws
- Taxes and tariffs that discourage energy efficiency activities
- Poor infrastructure
- Customs and maintaining the status quo.

CHALLENGES AND REWARDS OF IMPLEMENTATION OF ENERGY EFFICIENCY IN DEVELOPING COUNTRIES: RESIDENTIAL, COMMERCIAL, AND INDUSTRIAL SECTOR CASE STUDIES AND EXAMPLES

How Economic Development Is Measured

The statistical information that indicates the level of economic development can be used in the DSM decision-making process. Economic development can be measured through a number of social, economic, and demographic indexes (indicators). The following indicators are often used to measure the economic development of a country:

- *Per capita income*. This is a statistic that is seldom readily available, and GNP or GDP per capita is often used instead. According to 1999 GNP per capita, calculated using the World Bank Atlas method,[9] the following grouping of countries is possible: low-income, $755 or less; lower-middle-income, $756–$2995; upper-middle-income, $2996–$9265; and high-income, $9266 or more.
- *Per capita purchasing power*. This is a more meaningful measure of actual income per person, because it includes not only income, but also the price of goods in a country. It is usually measured in per capita GDP. Per capita GDP can be expressed both by using the current exchange rates in USD and by using the current purchasing power parities (PPPs). Purchasing power parities are the number of currency units required to buy goods equivalent to what can be bought with 1 unit of the currency of the base country or with 1 unit of the common currency of a group of countries.
- *Economic structure of the labor force*. In the United States and Western Europe, fewer than 2% of the workers are engaged in agriculture, whereas in certain African nations, India, and China, more than 70% of the laborers are in this sector. More than 75% of U.S. laborers are engaged in wholesaling, retailing, professional and personal services (including medical, legal, and entertainment), and information processing (such as finance, insurance, real estate, and computer-related fields).[10]
- *Consumer goods purchased*. The quantity and quality of consumer goods purchased and distributed in a society

also provide a good measure of the level of economic development in that society. Televisions, automobiles, home electronics, jewelry, watches, refrigerators, and washing machines are some of the major consumer goods produced worldwide on varying scales. The ratio of people to television sets in developing countries is 150 to 1, for example, and the ratio of the population to automobiles is 400 to 1. In California, the ratio is almost 1 to 1 for these consumer items.[10] The number of consumer goods such as telephones and televisions per capita is a good indicator of a country's level of economic development.

- *Education and literacy of a population*. The more men and women who attend school, usually the higher the level of economic development in a country. The literacy rate of a country is the percentage of people in the society who can read and write.
- *Health and welfare of a population*. Measures of health and welfare, in general, are much higher in developed nations than in les- developed ones. One measure of health and welfare is diet. Most people in Africa do not receive the United Nations' daily recommended allowance of diet. People in less-developed countries also have poor access to doctors, hospitals, and medical specialists.
- *Per capita energy consumption*. The energy consumption of the population is a good indicator of a country's level of economic development.

Transfer of Energy Efficiency Technologies

There may be more indicators of economic development to list, but what is sought for purposes of making energy policy is (preferably) a list of readily available indicators for deciding on the feasibility of the transfer of energy efficiency technologies. In a previous study,[9,11] a similar analysis was carried out for DSM technology transfer to developing countries.

Indicators such as GDP, per capita energy consumption, installed capacity, capacity factor, growth rate, income distribution, and carbon emissions will only give an idea about the level of the economic development of the country in question. More information is needed for choosing the correct DSM options. Details on the characteristics of end uses, which are seldom available in developing countries, would be a very useful indicator. These data can be obtained by monitoring, auditing, or surveying. In a recent study,[9] it was found that surveying is the least time-consuming low-cost method to obtain such data. To generalize the method, the indicators can be grouped under two headings: macro- and micro-level indicators (see Fig. 3).

The macro indicators are those that will be useful in determining which of the following categories the developing country belongs to. The categories are as follows:

- Advanced developing countries—middle-income countries that are advanced industrially
- Developing countries—low-income and lower-middle-income countries that are developing
- Least developed countries—low-income countries that are either not developing or developing very slowly.

The determination of which category a country belongs to is based on a cross-check of macro- and micro-level indicators. The execution of the decision-making is carried

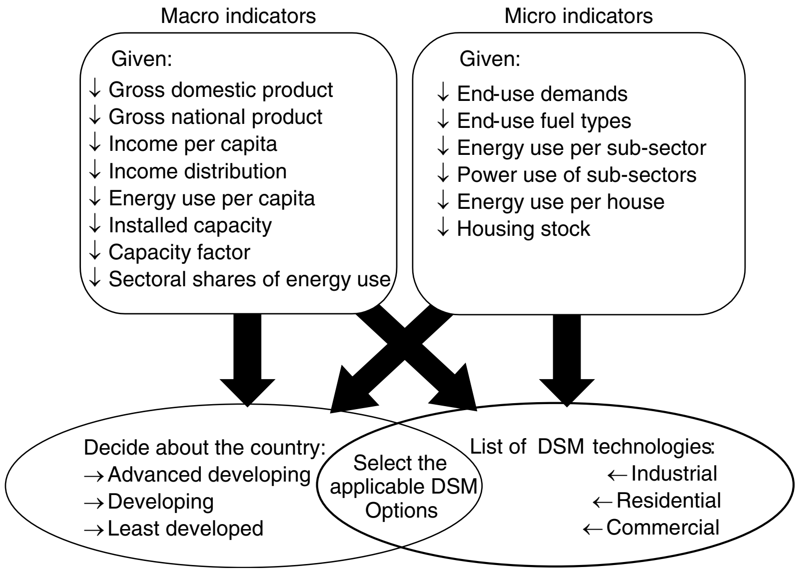


Fig. 3 Macro- and micro-level input variables used in decision-making for the most applicable DSM options in any developing country. Source: From Eastern Mediterranean University and Elsevier Science (see Refs. 9 and 11).

out through the consideration of the variables shown in Fig. 3. Using the end-use information, the available DSM options can be determined from the pool of DSM technologies. Then the applicability of the DSM technologies to the country in question is examined further by using information about the category of the country. The final decision is made by checking the cost-to-benefit ratio of the applicable DSM program.

Finally, the implementation of the DSM programs is evaluated, and if there are any doubts about their success, more surveys may be conducted to update the data at the micro level.

Reducing Power Capacity Needs

In developing countries, the slow-occurring technology transfer and, hence, delayed energy efficiency improvements in all sectors may lead to exaggerated power demand. In Turkey, the growth of installed power capacity has been greater than that of the GNP projected between 1968 and 2003 (see Fig. 4). In 1997, installed power was almost three times greater than it should have been with the GNP-projected levels; and in 2003, installed power was almost four-and-a-half times greater than those levels. The gap between the two trends (shaded area in Fig. 4) is a common feature of the developing countries. It is the reverse of JAWS, which was described for developed countries in the introduction. To close the gap between the two trends, in Fig. 4, the old infrastructures that are associated with the energy services should be renewed both at the power supply side and at the demand side. Therefore, this gap is referred to as an excess energy consumption (EXEC) gap.

The traditional method of meeting the demand with little attempt to rebuild the infrastructure and increase the efficiency of energy services needs to be revised in developing countries. Traditionally, the policy-makers in developing countries believed that the opportunity in power reduction lay in reducing the EXEC. The opportunity for power reduction is not limited to this amount, however. As the replacement of the old infrastructure continues, ways of improving energy efficiency at customer level and, hence, further reducing the power capacity requirements at the utilities and achieving the JAWS, a feature of developed countries, should also be considered. The gap between the maximum installed power and the minimum that can be achieved is called the energy-saving opportunity gap (ESOG).[12] The energy-saving opportunity in developing countries lies in reducing both EXEC and JAWS (i.e., ideally ESOG = EXEC + JAWS).

Case Studies of DSM Implementation

The two countries examined were Northern Cyprus and Turkey.[11] The flow chart in Fig. 5 shows a simplified approach that was applied for determining the best applicable DSM technologies for Northern Cyprus and Turkey. In this simple approach, the GNP per capita income of the country in question was used to determine the income level of the country. At the same time, a decision was made about the industrial development level of the country. For this reason, the shares of electricity consumption were used. This synthesis was carried out at the macro level, leading to setting priorities on considering

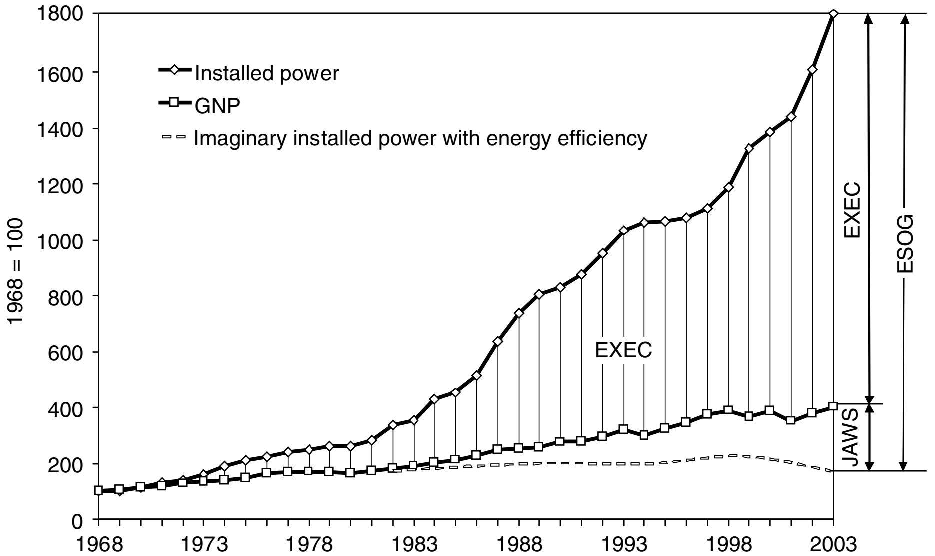


Fig. 4 Economic and power capacity developments of Turkey between 1968 and 2003.
Source: From Prime Ministry of Republic of Turkey, State Institute of Statistics Publication (see Ref. 13).

Energy Eff–Serv

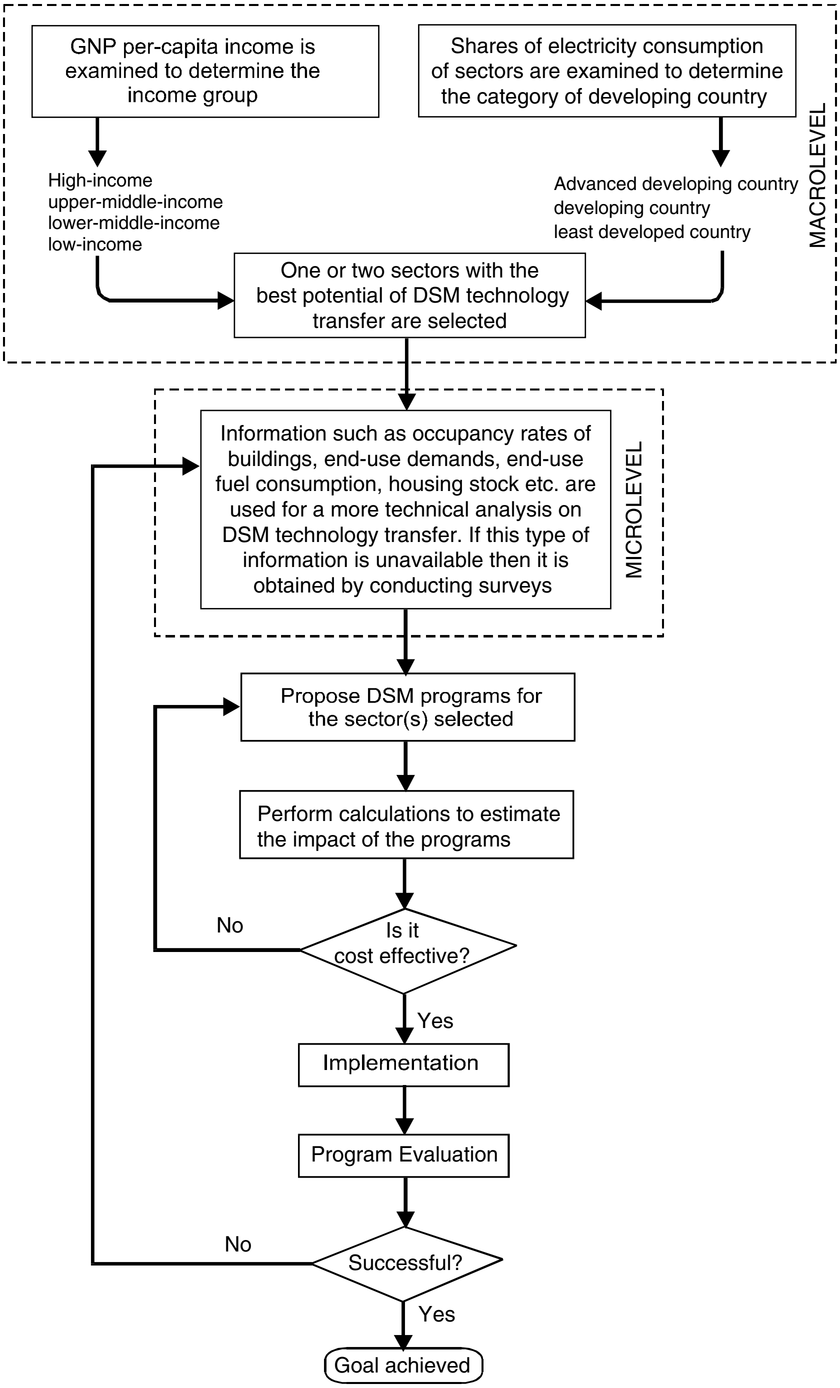


Fig. 5 A simplified approach for DSM technology transfer to developing countries. Source: From Elsevier (see Ref. 11).

the sectors with the best applicability of DSM options. To determine the specific DSM options to apply, however, detailed information on the end uses was required. Therefore, a survey was conducted for this purpose, and DSM programs were proposed.

Northern Cyprus, being a small island, is not an industrially advanced country. Sector-based electricity consumption shares are given as 35% for residential, 18.74% for commercial, and 9.39% for industrial sectors.[11] The great majority of the population (approximately 70%) is employed in the services sector; hence, the DSM measures are best selected for the residential and commercial technologies. The 1996 GNP per capita being $4222 USD, Northern Cyprus is an upper-middle-income country according to the World Bank criterion. The population is highly educated, and the literacy rate is almost 100%. With these macro indicators, it was decided that the residential and commercial sectors would have the best potential for load reduction and that the DSM technology transfer should be considered for these sectors. Surveys were conducted in both sectors to determine the micro indicators leading to the selection of the DSM options to apply.

In the residential sector, statistical information was obtained on the number of electrical appliances and their time of use, and end-use load curves were obtained. It was discovered that electrical water heaters demanded 50 MW at the winter peak hour (19:00), which constituted 45% of the 110-MW peak. The other shares of end-use demands of the winter peak were estimated to be 28% for space heating, 9% for television sets, 7% for lighting, and 5% for refrigeration. Demand-side management programs concerning the water heating, space heating, and lighting activities were proposed. The programs proposed were estimated to defer the need for a new 60 MW power unit worth $100 million (U.S. dollars) for at least 19 years at an expense of $12 million (U.S. dollars).[14]

The diversity of the commercial buildings and their end uses in Northern Cyprus preclude the adoption of the approach used in the residential sector. Not only do the system sizes for end uses vary greatly, but also, their time of use may vary for different building types. The problem can be simplified by selecting a segment of the commercial sector. In Northern Cyprus, the tourism sector was selected, being the segment in the commercial sector that consumes the most electricity. According to the surveys conducted, the power demand of this sector is greater during the summer season due to the heavy use of air conditioners. Air-conditioner use in the hotels constituted 27% of the utility load in summer. The proposed DSM technology transfer to this sector could reduce the summer peak by 11% at a total cost of $5,60,000 (U.S. dollars).[15]

Turkey, on the other hand, is an advanced developing country due to its past two decades of industrial development. The primary energy consumption share of its industrial sector is currently 29% and is expected to reach 40% in 2020. Due to its nonuniform income distribution, however, the 1999 GDP per capita is $2807 USD, which makes Turkey a lower-middle-income country. The level of education is low, and a rebate program on replacing incandescent lamps with compact fluorescent lamps (CFLs) in the residential sector is not expected to be as successful as in Northern Cyprus, especially in the low-income homes. The macro indicators of Turkey imply that the transfer of DSM technology would be best selected from the industrial options. Micro indicators such as end-use fuel types, average grid losses at 20%, and the magnitude of power and rate of heat demands of the factories pointed to sizable energy gains from a cogeneration program. A more detailed study[16] showed that by replacing the nuclear program with the cogeneration program, Turkey not only would save $72.6 billion (U.S. dollars), but also would reduce total primary energy demand by 11% in 2020.

CONCLUSIVE COMMENTS

Statistics and projections show that not only the rate of energy consumption, but also the carbon emissions of developing countries are rising very fast. In 2001, carbon dioxide emissions from industrialized countries were 49% of the global total, followed by developing countries at 38% and Eastern Europe/Former Soviet Union (EE/FSU) at 13%. If business continues as usual, by 2025, developing countries are projected to account for the largest share of world carbon dioxide emissions, at 46%, followed by the industrialized world at 42% and the EE/FSU at 12%.[17] To save the world, it is essential not only to transfer the latest energy efficiency technologies to the developing countries to achieve substantial reductions in energy consumption, but also to explore ways of rebuilding their poor infrastructure.

On the other hand, energy efficiency technologies, which are successfully used in many industrialized countries, may not always be feasible to apply in developing countries. Some DSM technologies may be easily transferable to some developing countries but not applicable in others.

Compact fluorescent lamps, for example, which were offered as a DSM measure in some developing countries, may not be suitable for other developing countries. In Turkey, the squatters, who are low-income people, probably would find this program luxurious, and even if they were given a full rebate, they might prefer selling the lamps rather than using them for themselves. On the other hand, in the villages of Bangladesh, 90% of the energy required for lighting is supplied from petroleum products; only 10% comes from electricity. Therefore, CFLs cannot be worthy of any consideration as a DSM option.

It is possible that some countries, although having a lower per capita income than others, may be industrially advanced, and the DSM technologies devised for the industrial sector may be more suitable than those for the residential sector. Turkey, for example, falls into the category of "industrially advanced developing country," and applying DSM measures in the industrial sector would be more successful and have more impact on the utility load management of the country.

Statistical information on end uses is also very useful in determining the suitability of the proposed energy efficiency or DSM technology. In Northern Cyprus, it was discovered that the largest demand for electricity came from domestic water heating. In South Africa, figures as high as 40%–50% of the monthly electricity use of an average middle-to-upper-income household were quoted for water heating. DSM technologies regarding efficient water heating would be most appropriate for these countries, but in many other countries, this may not be the case. In Thailand, for example, the highest potential savings could come from improving the energy efficiency of the refrigerators, the reason being the intense use of old-technology, inefficient refrigerators.

Therefore, it can be concluded that to apply energy efficiency technology transfer to developing countries, the social and economic realities of these countries need to be taken into account, along with the technical applicability of the programs. The approach should be twofold: First, the economic and social state of the country should be determined; and second, the suitability of the energy efficiency options with the highest potential savings should be determined.

REFERENCES

1. Rosenfeld, A.H.; Ward, E. Energy use in buildings. In *The Energy-Environment Connection*; Hollander, J.M., Ed., Island Press: Washington, DC, 1992; 223–257.
2. Levine, M.D.; Koomey, J.G.; Price, L.; Geller, H.; Nadel, S. Electricity end-use efficiency: Experience with technologies, markets and policies throughout the world. Energy Int. J. **1995**, *20* (1), 37–61.
3. Energy Information Administration *International Energy Outlook*; U.S. Department of Energy: Washington, DC, 2000.
4. U.S. Bureau of Census *International Database*; U.S. Bureau of Census: Washington, DC, 2001.
5. United Nations *1997 Energy Statistics Yearbook*; United Nations: New York, NY, 1998.
6. Office of Technology Assessment. *Energy in Developing Countries*, U.S. Congress Publication, OTA-E-486. U.S. Congress: Washington, DC, 1991.
7. Office of Technology Assessment. *Fueling Development—Energy Technologies for Developing Countries*. U.S. Congress Publication, OTA-E-516. U.S. Congress: Washington, DC, 1992.
8. Painuly, J.P.; Reddy, B.S. Electricity conservation programs: Barriers to their implementation. Energy Sources **1996**, *18* (3), 257–267.
9. Atikol, U. Utility demand-side management technology-transfer and impact assessment methodologies for developing countries [PhD thesis]. Eastern Mediterranean University: Magusa, Northern Cyprus, 2001.
10. Stutz, F.P.; Souza, A.R. *The World Economy: Resources, Location, Trade and Development*, 3rd Ed.; Prentice Hall: Englewood Cliffs, NJ, 1998.
11. Atikol, U.; Guven, H. Feasibility of DSM-technology transfer to developing countries. Appl. Energy **2003**, *76* (1–3), 197–210.
12. Atikol, U. An algorithm for reducing the power capacity needs in developing countries. In: *Modelling in Energy Economics and Policy*. Sixth IAEE European Conference, Zurich, Switzerland, September 1–3, 2004; International Association for Energy Economics, Cleveland, Ohio, 2004.
13. State Institute of Statistics *Statistical Year Book 2004*; Prime Ministry of Republic of Turkey, State Institute of Statistics Publication: Ankara, Turkey, 2005.
14. Atikol, U.; Dagbasi, M.; Guven, H. Identification of residential end-use loads for demand-side planning in northern cyprus. Energy Int. J. **1999**, *24* (3), 231–238.
15. Atikol, U. A demand-side planning approach for the commercial sector of developing countries. Energy Int. J. **2004**, *29* (2), 257–266.
16. Atikol, U.; Guven, H. Impact of cogeneration on integrated resource planning of Turkey. Energy Int. J. **2003**, *28* (12), 1259–1277.
17. Energy Information Administration *International Energy Outlook*; U.S. Department of Energy: Washington, DC, 2004.

Energy Efficiency: Information Sources for New and Emerging Technologies

Steven A. Parker
Pacific Northwest National Laboratory, Richland, Washington, U.S.A.

Abstract
The purpose of this entry is to share a list of useful organizations that provide reliable information on new and emerging energy-efficient technologies based on research and experience. Experienced energy managers may use the information provided by these organizations to enhance their knowledge and understanding, thereby improving their energy management programs. The scope is limited to publicly available and open-membership organizations that deal with new and emerging energy-efficient technologies, strategies, and products. The sources identified should not be considered exhaustive, but rather a first step "go to" list suggested by the author when searching for useful information on new and emerging energy-efficient technologies.

BACKGROUND

There are many strategies that can be used to reduce energy consumption and costs. These strategies might include:

- Awareness programs to motivate users and occupants to change behaviors that result in saved energy;
- Operations and maintenance programs, including commissioning to keep equipment running efficiently and to optimize run times;
- Training programs to educate people on how to reduce energy consumption and costs;
- Equipment procurement programs to buy the most efficient equipment available (rather than the cheapest first cost) to minimize life-cycle costs, which include energy costs;
- Commodity procurement programs to buy the least expensive energy source available; and
- Energy-efficient technology programs to replace existing or conventional equipment with more efficient equipment. This strategy also includes new and emerging energy-efficient technologies.

An effective energy management program will likely use many, if not all, of these strategies.

Removing old, inefficient equipment and replacing it with newer, more energy-efficient equipment is one method of reducing a site's energy consumption and costs. Two questions that are often asked include: "Which energy-efficient technology should I select?" And, just as important, "From what list of alternatives should I consider?"

The answer to the first question should be "I should select the alternative with the lowest life-cycle cost." The answer to the second question may be more difficult. If you do not place a new technology on the list of alternatives, it can never be selected. One issue with new and emerging technologies is the concern about risk and reliability. When something is new, it is perceived to have more risk. "It claims to be more energy efficient, but I have not seen it in operation. Does it really work? Will it work here? Will it work in this application? Is my maintenance team ready to work with this new technology?" These are all important and relevant questions.

The time it takes for a new technology to be accepted to the point of being considered conventional can be surprising. Take, for example, the fluorescent T-8 lamp and electronic ballast. How many years did it take for this technology to be seen as conventional for standard office lighting configurations and to fully penetrate the market? Think of the energy savings that could have been realized if this technology would have fully penetrated the market in only 5–10 years instead of 15–20 years. The same is true for all new equipment end uses (cooling, heating, ventilation, water heating, battery chargers, etc.). New technologies take a long time to fully penetrate the market and considerable energy may be saved if this transitory period were reduced.

Of course, balanced against the potential energy reduction is the potential for risk. Not every new technology will be cost effective and survive the market process. The way an experienced energy manager manages the risk, or at least better understands the risk, is to evaluate the new technology on a limited basis before it is mass deployed. This is the purpose of research, evaluations, demonstrations, pilot projects, and case

Keywords: New technology; Energy-efficient technology; Emerging energy-efficient technology; Energy efficiency; Energy-efficient equipment; State-of-the-art technology.

studies. With these actions, experienced energy managers can look before they leap, and in some cases, benefit from the lessons learned from others.

INTRODUCTION

Finding useful and reliable information on new and emerging energy-efficient technologies can be daunting, even for the experienced energy manager. Today, more than ever, a lot of information is available. There are books, professional associations with journals, magazines, e-newsletters, other periodicals, and the gamut of Internet Web sites. A lot of the information available is even well intended. One of the problems is that there is too much information to sort through. Finding the right type of information on the right subject can be difficult.

There are several reasons why it is difficult to find good information on new and emerging energy-efficient technologies. First and foremost, there may be little information on the technology and its potential effectiveness because the product is new. Second, collecting, analyzing, and reporting information on technologies or equipment can be expensive and time consuming. Third, many tests may be performed in-house, and may not be publicly available. Even with these barriers, there is a lot of information out there. In addition to their regular, full-time duties, energy managers can spend years building up a network of sources and resources.

There are a number of periodicals that can be used as sources for information on new and emerging energy-efficient technologies. In addition to journals provided through membership organizations, there are a number of low-cost and no-cost periodicals, many of which can be very useful in learning about new technologies or new approaches to energy efficiency. Unfortunately, there are so many periodicals that it may be impossible for any energy manager to stay on top of their reading list.

The purpose of this entry is to share a list of useful sources. This is not a list of sources on energy management—that list would be too long—but a list of sources (organizations) that specifically deal with new and emerging, energy-efficient technologies, strategies, and products. The sources identified in this entry should not be considered exhaustive. In fact, to keep its length manageable, this list has purposely been limited to sources within the United States. The author acknowledges that there are many organizations outside the United States performing similar useful services and many organizations, both public and private, that perform a lot of good research and provide a host of useful services in the area of energy efficiency. The list of sources included in this article may be considered an initial or first step or a go-to list suggested by the author in your search for useful information on new and emerging energy-efficient technologies.

Sources are not one-size-fits-all. Not every source does everything. The reader will find that some sources are more research- and development-oriented and focus on pre-commercial technologies, while others are more focused on enhancing the deployment of commercially-available technologies. Some sources are research-oriented, while others are more applied. In addition, some sources simply document results, whereas other sources document more detailed analysis, expand on the lessons learned, or offer guidance and tools to support future applications. You may also find different definitions for the terms "new" and "emerging," as well as "technology."

It is the author's hope that the reader will benefit from the information made available on these proactive organizations. Furthermore, it is the author's hope that more organizations will take proactive steps and make even more (reliable) information on new and emerging energy-efficient technologies publicly available.

UNITED STATES GOVERNMENT AGENCIES AND PROGRAMS

While some believe we should be spending more, the United States federal government supports a lot of research and development in the area of energy efficiency and renewable energy. Because the federal government supports a lot of this work, it is also a good source for information on new and emerging energy-efficient technologies. The information, however, can be difficult to find. This section identifies several sources of information within the federal government.

U.S. Department of Energy

Obviously, one of the largest sources of information on energy-efficient technologies is the U.S. Department of Energy (DOE). Within the DOE, there are three major offices that the advanced energy manager should be aware of—the Office of Energy Efficiency and Renewable Energy (EERE), the Office of Fossil Energy (FE), and the Office of Electricity Delivery and Energy Reliability (OE).

Each of the DOE program offices plays a unique role in the larger energy picture. FE deals primarily with power generation; OE deals primarily with electric delivery and infrastructure; and EERE deals primarily with end use. However, you will find some overlap among the DOE program offices. For example, distributed generation (renewable energy, fuel cells, cogeneration, etc.) is one of those technology sets that can be found in several programs.

Energy Efficiency and Renewable Energy

Simply put, the EERE mission is to strengthen America's energy security, environmental quality, and economic vitality in public–private partnerships that:

- Enhance energy efficiency and productivity;
- Bring clean, reliable, and affordable energy technologies to the marketplace; and
- Make a difference in the everyday lives of Americans by enhancing their energy choices and quality of life.

EERE manages and supports a wide variety of energy research, development, and deployment activities. EERE (www.eere.energy.gov) consists of 10 major program offices:

- Biomass Program (www.eere.energy.gov/biomass)
- Building Technologies Program (www.eere.energy.gov/buildings)
- Federal Energy Management Program (www.eere.energy.gov/femp)
- FreedomCAR and Vehicle Technologies Program (www.eere.energy.gov/vehiclesandfuels)
- Geothermal Technologies Program (www.eere.energy.gov/geothermal)
- Hydrogen, Fuel Cells, and Infrastructure Technologies Program (www.eere.energy.gov/hydrogenandfuelcells)
- Industrial Technologies Program (www.eere.energy.gov/industry)
- Solar Energy Technologies Program (www.eere.energy.gov/solar)
- Weatherization and Intergovernmental Program (www.eere.energy.gov/wip)
- Wind and Hydropower Technologies Program (www.eere.energy.gov/windandhydro)

Each of these program offices publishes or supports the publication of several reports every year. Many of these publications are available through the DOE Web site. Because the Web site is rather large, use the search engine.

While a large part of DOE is actively involved in the research and development of new, energy-efficient technologies, there are three specific subprograms that the author would like to highlight: the Industrial Technologies Program's inventions and innovations, Building Technologies Program's emerging technologies, and the Federal Energy Management Program's new technologies.

Inventions and Innovation (I&I), part of the DOE Industrial Technologies Program, provides grants to independent inventors and small companies with sound ideas for energy-efficient technologies. Distinct from other federal grant programs, I&I provides grantees not only with funding, but with additional resources such as training, market assessments, technical assistance, access to promotional events and materials, and special contacts to aid in commercialization endeavors. Although I&I is part of the Industrial Technologies Program organization, their mission is cross cutting; I&I supports energy efficiency and renewable energy technology development in focus areas that align with 10 EERE programs. For more information on I&I, see www.eere.energy.gov/inventions.

Emerging Technologies, part of the DOE Building Technologies Program (BTP), advances the research and development of the next generation of energy-efficient components, materials, and equipment. To help near-term emerging technologies overcome market introduction barriers, the program uses a combination of strategies, including demonstration/evaluation and technology procurement. These strategies are designed to increase buyer confidence and build demand for new energy-efficient technologies. For more information on BTP's Emerging Technologies, see www.eere.energy.gov/buildings/emergingtech.

The DOE Federal Energy Management Program (FEMP) supports a limited number of technology demonstrations through their new technology activities, which are designed to provide independent performance data to federal decision-makers and support timely federal adoption of energy saving and environmentally beneficial technologies. While these publications are developed to benefit the federal energy and building manager, they are useful to any energy manager wishing to advance their energy management program. For more information on FEMP's new technology publications, see www.eere.energy.gov/femp/technologies/tech_demos.cfm.

Office of Fossil Energy

According to the FE Web site (www.fossil.energy.gov), the primary mission of DOE's Office of Fossil Energy is to ensure that we can continue to rely on clean, affordable energy from our traditional fuel resources. Fossil fuels supply 85% of the nation's energy, and FE is working on such priority projects as pollution-free coal plants, carbon sequestration, FutureGen, more efficient power generation, more productive oil and gas fields, and the continuing readiness of federal emergency oil stockpiles.

While FE deals primarily with fossil fuels and source fuels for power generation, a significant amount of fuel-cell-related R&D is supported by this DOE office. If you are looking for information on fuel cells, do not overlook this important source.

Office of Electricity Delivery and Energy Reliability

According to the OE Web site (www.oe.energy.gov), their mission is to lead national efforts to modernize the electric grid, enhance the security and reliability of the energy infrastructure, and facilitate recovery from disruptions to the energy supply. OE is a relatively new office within the DOE that was formed when the Office of Energy Assurance and the Office of Electric Transmission and Distribution combined. Research and development programs within OE include: security and reliability of the transmission infrastructure, distributed energy, transmission grid modernization, energy storage, and superconductivity. New technologies related to the utility grid, including grid reliability and GridWise, are supported through this office. In addition, the distributed energy programs support distributed generation, combined heat and power, and thermally activated technologies.

DOE National Laboratories

To accomplish its mission, the DOE maintains several national laboratories and technology centers (www.energy.gov/organization/labs-techcenters.htm). Of the 13 DOE national laboratories, 11 are actively involved in EERE-supported research, development, and deployment activities (www.eere.energy.gov/site_administration/doe_labs.html). All publicly available reports published by DOE national laboratories are available through the National Technical Information Service (NTIS) at www.ntis.gov/search.

Each DOE national laboratory has specialty or focused areas of research, although there is considerable overlap in some areas. While the DOE national laboratories may be performing world-class research, their Web sites may not fully identify the breadth or depth of their knowledge and the information available. This is where nothing beats a good professional network or a (series of) well placed telephone calls.

Office of Scientific and Technical Information

At the DOE Office of Scientific and Technical Information (OSTI) (www.osti.gov), you can search for DOE research results and find out about ongoing research projects. OSTI makes R&D findings available to DOE researchers and the public.

U.S. Environmental Protection Agency

Under the U.S. Environmental Protection Agency (EPA), the Office of Air and Radiation, and the Office of Atmospheric Programs (OAP) is the Climate Protection Partnership Division (CPPD). The CPPD (www.epa.gov/cpd) promotes greater use of energy-efficient products and practices in the residential, commercial, and industrial sectors through the ENERGY STAR® partnership program (www.energystar.gov). The Division's programs, such as Climate Leaders (www.epa.gov/climateleaders), provide information, technical assistance, and recognition for environmental leadership to organizations as they develop strategies and take advantage of proven opportunities to reduce their greenhouse gas emissions. CPPD also encourages investments in efficient, clean technologies, such as combined heat and power (CHP) (www.epa.gov/CHP) and green power (www.epa.gov/greenpower) from renewable resources.

ENERGY STAR is a joint program of the EPA and the DOE. In 1992, the EPA introduced ENERGY STAR as a voluntary labeling program designed to identify and promote energy-efficient products to reduce greenhouse gas emissions. The ENERGY STAR label is now on major appliances, office equipment, lighting, home electronics, and more. EPA has also extended the label to cover new homes and commercial and industrial buildings.

[Note from the author: While I would trust the information on the ENERGY STAR Web site (www.energystar.gov) and information published by EPA and DOE; I would not trust any product or claim just because it had an ENERGY STAR label. ENERGY STAR is a voluntary program with a marketing foundation, and it is not difficult to find their logo being utilized by others to make unsubstantiated claims.]

U.S. Department of Commerce

The Department of Commerce is a very diverse organization within the federal government. Within the Department of Commerce, Office of Technology Administration, however, are two specific organizations an advanced energy manager should be aware of. These are the National Technical Information Service (better known as NTIS) and the National Institute of Standards and Technology (better known as NIST).

National Technical Information Service

The National Technical Information Service (NTIS) is part of the U.S. Department of Commerce's Technology Administration. NTIS' primary purpose is to assist U.S. industries to accelerate the development of new products and processes, as well as helping the United States maintain a leading worldwide economic competitive position. NTIS is the largest central resource for government-funded scientific, technical, engineering, and business-related information available today. They have information on more than 600,000 information products covering over 350 subject areas from over 200 federal agencies. To search the NTIS library for federal reports and documents published since 1990, go to www.ntis.gov/search/

National Institute of Standards and Technology

The National Institute of Standards and Technology (NIST) is a nonregulatory federal agency within the U.S. Commerce Department's Technology Administration. NIST's mission is to develop and promote measurement, standards, and technologies to enhance productivity, facilitate trade, and improve the quality of life. NIST consists of four primary programs. Of specific interest to energy efficiency are the NIST laboratories (www.nist.gov/public_affairs/labs2.htm) and the

Advanced Technology Program (www.atp.nist.gov). The NIST laboratories conduct research that advances the nation's technology infrastructure and is needed by U.S. industry to continually improve products and services. The Advanced Technology Program accelerates the development of innovative technologies for broad national benefit by cofunding R&D partnerships with the private sector.

Other Federal Agency Sources

As a result of legislation, every federal agency has an internal energy management organization and/or program working toward a series of energy-efficiency goals. In response to these demands, several federal agencies are performing in-house research on new and emerging energy-efficient technologies to identify those technologies that can best serve the goals of the agency. Some of this information is publicly available on agency Web sites (although it can be very difficult to find), while other agencies have found it necessary to keep the information on internal-access-only intranets.

Department of Defense, Fuel Cell Program

There are a number of energy-efficiency research and development programs within the U.S. Department of Defense (DOD), each serving specific DOD missions. While most of the energy-efficiency programs serve internal DOD objectives, the DOD fuel cell program (dodfuelcell.cecer.army.mil) has provided significant benefit to the overall advancement and commercialization of fuel cells. The DOD fuel cell program is managed by the U.S. Army, Corps of Engineers, Engineer Research and Development Center (ERDC), Construction Engineering Research Laboratories (CERL). The DOD fuel cell projects Web site provides information guides, demonstration results, and information on the climate change (fuel cell) rebate program.

STATE AGENCIES AND UNIVERSITY PROGRAMS

While the federal government has some very substantial research, development, and deployment programs for new and emerging energy-efficient technologies, several state agencies and university programs are also excellent resources and are often closer to technologies ready for commercial deployment.

While the resources are good, this presents a case where perspective, or bias, should be considered by the reader. State energy offices focus on the priorities of the state. Furthermore, climate factors can also skew results. Support infrastructure and product availability may also be regional. For example, a dehumidification technology demonstrated in Orlando will have significantly different findings in San Francisco. Similarly, sunny California is subject to both high electric energy and electric demand rates. Therefore, you may find considerable information supporting the technical appropriateness and cost effectiveness of a technology such as photovoltaic energy. The results will be significantly different for an application in the Pacific Northwest, where solar radiation has different technical factors and both energy and demand rates are notably lower. The author's point is that many factors can support the appropriateness of a new technology: end-use application, energy, peak demand, average demand, climate, location, and infrastructure are just some of these factors.

Almost every state in the United States has a state energy office. The experienced energy manager should become familiar with their local state energy office and the services it offers. In addition, there are several universities with energy-research programs. This section does not purport to be all inclusive; however, the following state and university programs are particularly noted for performing research and making publications available on new and emerging energy-efficient technologies.

California Energy Commission

The California Energy Commission (CEC) is one of the largest state energy programs. The CEC (www.energy.ca.gov) has program offices on energy efficiency, renewable energy, transportation energy, and significant research and development activities. Some CEC programs appear to be larger than their DOE counterparts.

Public Interest Energy Research

The California Energy Commission's Public Interest Energy Research (PIER) program creates state-wide environmental and economic benefits by supporting energy research, development, and demonstration (RD&D) projects that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy products and services to the marketplace. PIER includes a full range of RD&D activities that will advance science or technology not adequately provided by competitive and regulated markets. The PIER Web site is located at www.energy.ca.gov/pier.

While PIER is involved in RD&D, they are not involved—at least not directly—in the commercialization or deployment of technologies. In some cases, you may find that local utilities become involved with new technologies that achieve successful results from PIER demonstrations. For this reason, local utilities can be good sources for information. In California, the major utilities have collaborated to form the Emerging Technology Coordinating Council (ETCC). The ETCC is discussed in more detail later in this entry.

Energy Center of Wisconsin

The Energy Center of Wisconsin (www.ecw.org) is a private, nonprofit organization dedicated to improving energy sustainability, including support of energy efficiency, renewable energy, and environmental protection. The Energy Center of Wisconsin provides objective research, information, and education on energy issues to businesses, professionals, policymakers, and the public. The Energy Center's education department supports comprehensive energy education programs.

They offer hundreds of publications on a myriad of energy-efficiency topics and host a unique collection of industrial fact sheets and many other research materials. The Energy Center also has a comprehensive collection of energy-related library materials, many of which are available through their Web site.

Florida Solar Energy Center

The Florida Solar Energy Center (FSEC) is a state-supported renewable energy and energy-efficiency research, training, testing, and certification institute. FSEC (www.fsec.ucf.edu) is a research institute of the University of Central Florida (UCF) and functions as the state's energy research, training, and certification center. FSEC's mission is to research and develop energy technologies that enhance Florida's and the nation's economy and environment, and to educate the public, students, and practitioners on the results of the research. FSEC has gained national and international recognition for its wide range of research, education, training, and certification activities. FSEC maintains a wide variety of technical articles, research reports, newsletters, and public information documents. Their joint-use research library is reported to be one of the largest repositories of energy publications in the nation.

Iowa Energy Center

The Iowa Energy Center (www.energy.iastate.edu) invests its resources to create a stable energy future for the state of Iowa. The Iowa Energy Center works quietly and steadily, producing dividends that support Iowa communities, businesses, and individuals. Most of the Iowa Energy Center's activities are spent on energy-efficiency research, demonstrations, and education projects. These projects address energy use in agriculture, industry, commercial businesses, municipalities, and residential areas. The Iowa Energy Center is constantly at work seeking opportunities to improve the total energy picture of the state, its businesses, and communities.

The National Building Controls Information Program (www.buildingcontrols.org) was established by the Iowa Energy Center with support from the EPA to facilitate the adoption of energy-efficient building control products and strategies through testing, demonstration, education, and dissemination of product information. A special Web site, DDC Online (www.ddc-online.org), was developed by the Iowa Energy Center and is reported to be the most complete, unbiased listing available of direct-digital controls (DDC) vendors and information related to DDC systems.

New York State Energy Research and Development Authority

The New York State Energy Research and Development Authority (NYSERDA) is a public benefit corporation created by the New York state legislature. NYSERDA (www.nyserda.org) has been cited by the DOE as being among the best government research organizations in North America. NYSERDA's principal goal is to help all New York state utility customers solve their energy and environmental problems while developing new, innovative products and services that can be manufactured or commercialized by New York state firms, through which NYSERDA supports a wide assortment of research and development in energy-efficient technologies. Many of NYSERDA's technical reports are available through their Web site (www.nyserda.org/publications). In addition, many of NYSERDA's reports are available from NTIS.

Rensselaer Polytechnic Institute, Lighting Research Center

The Rensselaer Polytechnic Institute, Lighting Research Center (LRC) is a university-based research center devoted to lighting (www.lrc.rpi.edu). The LRC is a reliable source for objective information about lighting technologies, applications, and products. The LRC provides training programs for government agencies, utilities, contractors, lighting designers, and other lighting professionals. LRC programs cover a range of activities, including both laboratory testing of lighting products and real world demonstration and evaluation of lighting products and designs. They conduct research into energy efficiency, new products and technologies, lighting design, and human factors issues.

Washington State University, Energy Programs

The Washington State University (WSU) Energy Program (www.energy.wsu.edu) is a self-supported department within the university's extension service. The WSU Energy Program provides support to other national and local energy organizations, such as EnergyIdeas.org. The program has developed a significant library of materials on new and emerging technologies.

EnergyIdeas.org provides comprehensive information about commercial, industrial, and residential energy efficiency. You will find actual customer questions with detailed answers, case studies, reports, and articles on topics ranging from appliances and lighting to motors and solar energy. The EnergyIdeas Web site is sponsored by the nonprofit Northwest Energy Efficiency Alliance (www.nwalliance.org) and it is managed by the WSU Energy Program.

More on University Programs

There are several universities that operate energy management research programs. Each major university energy program will have specialty or focused areas of research. Unfortunately, there does not appear to be a composite index identifying the universities and their points of contact. Furthermore, while the university programs may be performing world-class research, their Web sites may not fully identify the breadth or depth of their knowledge. This is where nothing beats a good professional network or a well-placed telephone call.

MEMBERSHIP ORGANIZATIONS AND ASSOCIATIONS

There are a number of membership associations. Some membership organizations are associations of companies, while others allow individual membership. Both types of organizations add value to the field of energy management. In addition to supporting research, training, and publications, several of these organizations can also respond to special inquiries from members. The following is a nonexhaustive list of membership associations and other associations that may be useful to members in obtaining information on new and emerging energy-efficient technologies.

Alliance to Save Energy

The Alliance to Save Energy (ASE) (www.ase.org) is a nonprofit coalition of business, government, environmental, and consumer leaders. The Alliance to Save Energy supports energy efficiency as a cost-effective energy resource under existing market conditions and advocates energy-efficiency policies that minimize costs to society and individual consumers, and lessen greenhouse gas emissions and their impact on the global climate. To carry out its mission, the Alliance to Save Energy undertakes research, educational programs, and policy advocacy; designs and implements energy-efficiency projects; promotes technology development and deployment; and builds public–private partnerships in the United States and other countries. Although the ASE is very policy-oriented (and, therefore, very politically oriented), it does support several technical programs.

American Council for an Energy-Efficient Economy

The American Council for an Energy-Efficient Economy (ACEEE) (www.aceee.org) is a nonprofit organization dedicated to advancing energy efficiency as a means of promoting both economic prosperity and environmental protection. ACEEE conducts in-depth technical and policy assessments, organizes conferences, and publishes books and reports. While ACEEE is not a membership organization, they do send out notices of publications, conferences, and other activities. ACEEE sponsors several events, but they are probably most noted for the biennial Summer Study on Energy-Efficient Buildings. Recently, ACEEE has also been organizing a biennial Summit on Emerging Technologies in Energy Efficiency.

Consortium for Energy Efficiency

The Consortium for Energy Efficiency (CEE) (www.cee1.org) is a nonprofit, public benefits corporation that promotes energy-efficient products and services. CEE develops initiatives for its members to promote the manufacture and purchase of energy-efficient products and services. CEE's goal is to induce lasting structural and behavioral changes in the marketplace, resulting in the increased adoption of energy-efficient technologies.

E Source

E Source Companies LLC (www.esource.com) is a membership-based information services company. E Source information services provide member organizations with unbiased, independent analysis of retail energy markets, services, and technologies. According to their Web site, E Source serves as a high-value filter of the torrent of information on developments in the energy services marketplace, sorting through the hype and providing clients with concise strategic insights and in-depth technology assessments. The E Source core products and services are offered through an integrated package of membership benefits. Additional services for energy service providers, including multiclient studies and focused research services, may be purchased.

Professional and Trade Associations

There are a number of professional and trade membership associations. Professional and trade associations provide their members benefits such as journals, publications, and training, as well as a professional network. Some organizations also sponsor certification programs, conferences and trade shows, technical research, or other benefits. The following is a nonexhaustive list of professional and trade membership associations with significant ties to advancing the field of energy management through the research, development, deployment, assessment, and/or the use of new and emerging energy-efficient technologies.

- Association of Energy Engineers (AEE), www.aeecenter.org
- American Institute of Architects (AIA), www.aia.org
- American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE), www.ashrae.org
- American Society of Mechanical Engineers (ASME), www.asme.org
- California Commissioning Collaborative (CCC), www.cacx.org
- Fuel Cells 2000, www.fuelcells.org
- Gas Appliance Manufacturers Association (GAMA), www.gamanet.org
- Geothermal Heat Pump Consortium, www.geoexchange.org
- Geothermal Resources Council (GRC), www.geothermal.org
- Illuminating Engineering Society of North America (IESNA), www.iesna.org
- Institute of Electrical and Electronics Engineers (IEEE), www.ieee.org
- Institute of Industrial Engineers (IIE), www.iienet.org
- International Facility Management Association (IFMA), www.ifma.org
- International Ground-Source Heat Pump Association (IGSHPA), www.igshpa.okstate.edu
- National Council on Qualifications for the Lighting Professions (NCQLP), www.ncqlp.org

UTILITY ENERGY CENTERS AND RESEARCH PROGRAMS

Many if not most utilities perform research on new and emerging energy-efficient technologies. However, much of this research is designed to support in-house programs. On the other hand, some utilities work aggressively with their customers to research, demonstrate, and deploy these new technologies. Even if the information is not made publication-ready, the experienced energy manager will benefit from establishing an active and on-going relationship with local utility customer account-service representatives. Through the account representatives, the energy manager will learn of the utility's involvement in new and emerging technologies and can benefit from the lessons learned.

California Statewide Emerging Technology Program

The statewide Emerging Technologies (ET) program seeks to accelerate the introduction of "near market-ready" energy-efficiency innovations that are not widely adopted by utility customers in California. Four investor-owned utilities (Pacific Gas and Electric, San Diego Gas and Electric, Southern California Edison, and Southern California Gas) and the California Energy Commission (CEC) are working together cooperatively to pool resources and knowledge for project selection and results dissemination.

The ET program consists of two main components: (1) demonstration and information transfer and (2) participation in the ETCC. The demonstration component provides technology assessments and information to utility customers and industry, often in the form of technology demonstrations at customer facilities. The ETCC coordinates activities and information among the utilities. The ETCC also maintains a Web site listing projects and summary results in its database (www.etcc-ca.com).

The ETCC is charged with administrating California utility, ratepayer-funded programs for energy-related research and energy-efficient emerging technologies. The ETCC coordinates among its members to facilitate the application of energy-efficient emerging technologies that will transform the market and benefit California ratepayers.

The Emerging Technologies project database contains a compilation of technologies and applications that are part of either current or recent projects that council members have sponsored. The database also contains information about the technologies, their applications, and the resulting assessments from member projects.

Energy Design Resources

Energy Design Resources (www.energydesignresources.com) offers energy design tools, software, design guides, case studies, and more. Their goal is to educate architects, engineers, lighting designers, and developers about techniques and technologies, making it easier to design and build energy-efficient commercial and industrial buildings in California. Energy Design Resources is funded by California utility customers and administered by Pacific Gas and Electric Company, Sacramento Municipal Utility District, San Diego Gas and Electric, Southern California Edison, and Southern California Gas under the auspices of the California Public Utilities Commission.

Lighting Design Lab

The Lighting Design Lab (www.lightingdesignlab.com) works to transform the Northwest lighting market by promoting quality design and energy-efficient technologies. The Lighting Design Lab accomplishes its mission through education and training, consultations, technical assistance, and demonstrations. The Lighting Design Lab is sponsored by the Northwest Energy Efficiency Alliance, Seattle City Light, Puget Sound

Energy, Snohomish County Public Utility District, British Columbia Hydro, and Tacoma Power.

Sacramento Municipal Utility District, Customer Advanced Technologies Program

The Sacramento Municipal Utility District (SMUD) Customer Advanced Technologies (CAT) program is a research and development program designed to encourage customers to use and evaluate new or underutilized technologies. According to one presentation, CAT provides the following benefits: it helps customers sort fact from fiction, identify the most promising technologies through direct, first-hand experience, and avoid making major investments in technologies that do not work. Unlike many R&D programs, research is accomplished through implementing real world demonstration projects (instead of laboratory testing). Completed demonstration projects include lighting technologies, light emitting diodes (LEDs), building envelopes, heating ventilation, and air-conditioning (HVAC) systems, as well as a wide variety of other technologies. Reports describing the results for many of these projects are available through their Web site (www.smud.org/education/cat/index.html).

Southern California Edison, Energy Centers

Southern California Edison (SCE) operates two energy centers (www.sce.com/RebatesandSavings/EnergyCenters)—the Agricultural Technology Application Center (ATAC) and the Customer Technology Application Center (CTAC). Both facilities offer hands-on demonstrations of the latest state-of-the-art technologies, as well as workshops, classes, and interactive displays.

Southern California Gas Company, Energy Resource Center

The mission of the Energy Resource Center (ERC) is to serve as a one-stop "idea shop" where customers can find the most efficient, cost-effective, and environmentally sensitive solutions to their energy needs. The ERC showcases innovations in resource-efficient designs, materials, and equipment to help businesses make informed choices about energy consumption and conservation. For more information, see www.socalgas.com/business/resource_center/erc_home.shtml.

OTHER SOURCES OF INFORMATION

The following organizations are worthwhile sources that did not fit into any of the previous categories.

Centre for Analysis and Dissemination of Demonstrated Energy Technologies

The Centre for Analysis and Dissemination of Demonstrated Energy Technologies (CADDET) is an international information source that helps managers, engineers, architects, and researchers find out about renewable energy and energy-saving technologies that have worked in other countries. Along with its sister program, GREENTIE (www.greentie.org), CADDET ceased collecting new information at the end of March 2005. Nevertheless, the information currently remains available through the Web site (www.caddet.org). CADDET's objective was to enhance the exchange of information on new, cost-effective technologies that have been demonstrated in applications such as industry, buildings, transport, utilities, and agriculture. The information was not only collected and disseminated to a very wide audience; it was also analyzed to provide a better understanding of the benefits of the technologies.

Portland Energy Conservation, Inc.

Portland Energy Conservation, Inc. (PECI) considers itself passionate about energy efficiency. Their mission is to help everyone use energy more effectively. PECI assists clients in the promotion of energy-efficient practices and technologies that benefit both businesses and individual consumers. PECI helps clients deliver long-term energy savings by helping transform markets through education and incentive programs that build demand for more efficient products and services. PECI's Web-based library (www.peci.org) offers considerable information on commissioning, as well as operations and maintenance.

CAVEAT EMPTOR

It is important to remember that the Internet is a tool and not a source. Almost anyone has the capability to make information available through the Internet. Furthermore, not every article in a periodical or presentation at a conference undergoes rigorous peer review. Every source has an intent or purpose when making information available. It is important to understand the source to determine potential bias (or perspective) and the reliability of the information. Even if the intent is altruistic, the reader needs to be aware of the qualifications, or basis, of the source, as well as any potential bias.

While each of the organizations identified in this entry provide publications or other informational sources about new and emerging energy-efficient technologies, as with all things new, there are risks. Some risks are minimized by evaluating new technologies on a limited basis before they are deployed on a larger scale. Demonstrations and pilot projects allow users to "look before you leap" when applying new technologies.

Many of the organizations identified in this entry are doing their part to get information on new and emerging energy-efficient technologies into the hands of energy managers, design engineers, building owners, and others, and they are encouraging them to consider the new technologies so that they may save energy, reduce costs, reduce emissions, or achieve other objectives.

The reader should also understand that most of these organizations are not making guarantees and implying endorsements. For example, those who claim endorsements from federal agencies are misleading the public. Furthermore, publications provided by these organizations are not substitutes for sound engineering or due diligence on the part of the reader. These programs do strive to be accurate and responsible. Remember, their objectives are to help.

New and emerging technologies are also subject to change. It is important to note the date of the publication or other information source. Publications usually offer a snapshot of a technology at a given time. As time passes, technologies, costs, maintenance recommendations, and even manufacturers change. In many cases, this can be a good thing; manufacturers can make numerous improvements to the equipment as more is learned about the technology and its operation. However, there may be other changes. New manufacturers may develop a product line that may not have been in that business when the report was published. Other manufacturers, who were known at the time of publication, may relocate, merge, consolidate, drop the product line, or even go out of business. Readers must continue to do their homework.

Energy managers and facility staff need to be wary. *Caveat emptor*, let the buyer beware, definitely applies. Readers are encouraged to check the original sources and follow up with other users. Do not rely on third-hand sources (for example, a government report should be checked out from a government source and not from a second- or third-party distributor).

CONCLUSIONS

Energy management is a diverse and growing field. As a result, there are many more sources available to the energy manager beyond those identified in this article. The organizations identified in this article are simply the ones this author turns to first when seeking useful information about an emerging or unknown technology. Many organizations are involved in innovation, research, and development. Many other firms are involved in assessment, experimentation, validation, training, and other forms of deployment. The energy manager can benefit from each of these organizations' contributions.

Technologies are continuously evolving. Efficiencies are improving, controls are improving, and maintenance is being simplified. New and emerging energy-efficient technologies can be used to help the experienced energy manager achieve and surpass their goals. Trying something new may not be risk-free, but these programs are doing their part to assist experienced energy managers.

Disclaimer

The information contained in this article is based on the author's 25 meandering years in energy management. The information does not purport to be comprehensive or complete, but is provided in the hope that it may be of some use to the reader. Any source provided in the entry should not be interpreted as an endorsement, just as any source not provided in this entry should not be considered a negative endorsement. If you do not want to use the information contained in this entry in the spirit in which it is intended, then do not use it. The reader should also be aware that there is a difference between 25 years of experience and 1 year of experience 25 times over. The author is unsure where he fits in the spectrum.

ACKNOWLEDGMENTS

The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830.

Energy Efficiency: Low Cost Improvements

James Call
James Call Engineering, PLLC, Larchmont, New York, U.S.A.

Abstract
In most energy efficiency initiatives, one is blessed with a number of energy-saving projects, all competing for a company's limited corporate resources. The various projects are typically interrelated; one of them, if implemented, would impact the others. They have a wide range of implementation costs and related energy-saving impacts—that is, they have different payback periods. If the quickest payback projects are done first, their savings will be unavailable to "subsidize" long payback projects. Thus, the long payback projects will have to fend for themselves in their fight for capital budget allocations. The end result may free up capital for other company projects in completely different areas.

This is why low-cost/no-cost energy projects are so valuable: they not only save money in their own right, but also set up a level playing field for the ongoing prioritization of corporate resources generally. Fourteen low-cost/no-cost projects from actual field experience are described as examples of the kind of projects that can introduce big savings up front and, thus, change the business case for long payback projects subsequently considered.

BACKGROUND

Longer-Term Consequences of Implementing Low-Cost Options First

If low-cost, quick-payback solutions are bundled in with longer-payback projects, the quick-payback projects can inadvertently subsidize the longer-payback projects that overlap on some of the benefits.

Consider a case in which an $1800, 12-channel "smart" time clock (holidays, sun tracker, outside air temperature sensor, etc.) would reduce energy costs by 25% and have a 2-mo payback. Alternatively, an $80,000, full-featured Energy Management System (EMS), if implemented alone, would reduce energy costs by 30% and have a 5-year payback. The issue is that if the 12-channel smart time clock is installed first, it co-opts the "savings" from the EMS project, stretching the EMS payback to perhaps 20 years. This is because the EMS also has a smart time clock inside it, as one of its many features, and that single feature was mostly why the EMS seemed attractive in the first place.

It generally will be impossible to develop a fair and accurate business case for a longer-payback system if low-cost, quick-payback solutions are "reserved" to bundle in with them.

Put another way, mixing in low-cost, quick-payback approaches to sweeten the case for longer-payback projects is analogous to improving a $5 bottle of wine by mixing it with a $90 bottle—a strategy perhaps to be avoided ambitiously.

Corporate Payback Analyses Should Include All Company Opportunities

In asking energy projects to compete for corporate resources (i.e., in comparing payback periods), companies will generally want to compare all projects, not just other energy projects. A proposed energy savings project, for example, should undergo comparison with projects in other departments, such as Advertising and New Product Development.

Maintaining or Improving Comfort and Safety Must be Addressed in All Analyses

In all projects, of course—low cost, no cost, or otherwise—employee comfort and safety, and compliance with applicable codes, must be attended to. Such requirements obviously should have the final say in all project decisions and implementations.

EXAMPLES FROM THE FIELD

Safety First

Any implementation project must put safety first. The engineering, design, and implementation should be done by qualified professionals only and with the greatest attention to safety.

Following are summaries of a number of low-cost/no-cost opportunities that have been encountered in actual

Keywords: ASHRAE 12; ASHRAE 62; Comfort and safety; Cost accounting feedback; Open loop; Payback period; Prioritize.

energy cost reduction initiatives. These are the kinds of opportunities that frame the issues in comparing payback periods and prioritizing competing projects.

Replace Lugs on Time Clocks to Reestablish Night Setback

Surprisingly, many buildings with installed time clocks have their OFF lugs removed and, thus, let the HVAC systems run 24/7. They were removed for reasons that probably seemed compelling, or at least expedient, at the time.

In a 100,000-ft^2. office building, for example, employees working late at night or on weekends had complained some years previously that the HVAC was not on and that it was too cold, or too hot. This was because the building was in the scheduled setback mode during off hours. As an expedient solution to this issue, the Building Services Department simply removed the OFF lugs on the 7-day time clock. Because energy expense was added in with numerous other overhead expenses and then charged back to about 20 departments on a square-foot basis, the sudden increase in energy costs went unnoticed for a number of years.

To reclaim the lost savings from the abandoned night and weekend setback, a low-cost solution was implemented in the following steps:

1. Replace the OFF lugs on the master time clock to run the HVAC 6:00 a.m. to 7:00 p.m. 5 days a week.
2. Install latching relays across the existing time clock's low-voltage outputs for the 12 building zones of interest, and wire their latch trigger circuits to user-accessible pushbuttons.
3. Provide a labeled "red button" in an accessible place (by the copy machine) in each of the 12 zones, and let employees know that if they come in to work during off hours, they should simply push the button, and their zone will come on for up to 6 h. Later, if the zone shuts OFF, employees should push the button again to turn on the zone for 6 more hours.
4. Provide a reset to the latch coil circuit by putting a $100 time clock in series with the latch current that turns OFF 1 min every 6 h.
5. Provide high-temperature and low-temperature "night setback" thermostats, which would ensure that the space did not get extremely hot in summer or so cold as to freeze pipes or fan coils in winter.

This solution cost under $1000, was implemented in 1 month, and saved more than $30,000 per year, giving it a nominal payback period of a couple of weeks.

Fig. 1 is a photo of the relay logic as actually constructed, and Fig. 2 shows the pushbutton setup that triggers the temporary comfort cycle.

Fig. 1 Relay logic to cycle certain zones, temporarily, with timed-out latch release.

Tighten Schedules When Possible

Building systems may be set on an automatic schedule that, upon review, can be substantially tightened without affecting employee comfort. A good way to determine the optimum schedule is to make a manual count of cars in the parking lot—say, every 15 min around the beginning and end of the business day. A rough hand-drawn plot of arrivals per quarter-hour period will indicate when the space should be ready to satisfy, say, 98% of the people. The optimum "be ready" time likely will not occur at an even quarter-hour (e.g., 7:30 a.m.), and if the existing schedule is in fact set at such even times (as most are), there is probably an opportunity to reduce energy usage.

Data loggers are invaluable in determining the warming and cooling curves of a building. The curves generally have the characteristic shape of the classical capacitor charge and discharge (i.e., an exponentially decaying rise or fall toward a final value).

A longer lead time may be needed on Monday mornings, of course, or any time when the building levels have drifted greatly during the off hours. Given the usual 30–90 min that most spaces need to achieve the target temperature, conditions would still be fairly comfortable, plus or minus 15 or 20 min of the actual target time.

Update Schedules to Reflect Current Building Use

During a 1:00 a.m. visit to a commercial office space as part of an energy audit, it was noticed that a large wing of the building was still running the HVAC system, even though the rest of the building had gone into night setback, as was expected, from the time clock.

Upon investigation, it was determined that the time clock did indeed have the zone still ON until 2:00 a.m. each weekday night, while all the other zones shut OFF at 7:00 p.m. It turned out that the zone still running formerly

Fig. 2 User pushbutton to start temporary comfort period.

was used by a second shift in a data entry department that finished at 2:00 a.m. That was no longer the case, but the time clock had never been reset when the late shift was discontinued. It had been more than 4 years since the late shift had used the space.

Because the cooling tower and the entire chiller plant were interlocked to run when any zone called for cooling, resetting the time clock to reflect the current schedule saved an estimated $8000 per summer.

Shut Off Boiler in Summer by Installing a Dedicated Domestic Hot Water Heater

Domestic hot water for a company's 10 bathrooms was provided by a loop in the building's main boiler running through a heat exchanger. Thus, the main boiler was kept on all summer, at a cost of some $25,000 per year for gas.

The solution was to install a 60-gal dedicated hot water tank for domestic hot water and to shut off the main boiler during the summer. The dedicated tank was fitted with a mixing valve, expansion tank, and pump to address ASHRAE 12[1] requirements.

The improvement took 2 months to implement, cost some $4000, and had a payback period of less than 3 months.

Reduce Gross System Overcapacity by Shutting off Equipment as Appropriate

It is not unusual to find gross overcapacity in HVAC systems. The reasons behind this appear to include

- "Rounding up" in each serial design step when the building was originally built. That is, installed overcapacity may have resulted from the cumulative effect of what the building owner initially saw as the load, then the design engineer's safety factor, and finally installing 7.50-ton units when the nominal calculation showed 6.74 ton.
- Building use changes since the original design.
- Accumulation of ad hoc package systems additions, over the years, to meet special needs.

The following two examples illustrate low-cost/no-cost approaches that were used to reduce overcapacity:

a. *Retiring a Package Unit and Attaching the Served Ductwork to a Nearby Remaining Unit*: A company's office space had three package air conditioners, each with dual 5-ton compressors. Theoretical calculations, which were confirmed by actual tests using loggers, determined that the installed cooling capacity was more than 3 times the required amount for cooling comfort on the hottest day of the year. Moreover, one of the units had been scheduled for replacement due to its age and unreliability.

To eliminate the need for the unit's replacement and reduce electric demand at the same time, the following steps were implemented:

1. Confirmed by field measurements that there would be sufficient air volume from the other two good units if the old unit were retired.
2. Decommissioned the old unit and relocated its discharge ductwork over to one of the adjacent remaining good units.
3. Rebalanced the served space.

The result was a savings of more than $12,000 of averted costs to replace the old unit, as well as a 20% drop in electric demand.

Energy Eff–Serv

b. *Run Only One Compressor in Two-Compressor Package Units*: A 90,000-ft^2. office building had 30 rooftop package units, each controlled by a local thermostat in the related zone. Calculations, confirmed by loggers, indicated that the building as a whole had well over twice the cooling capacity needed even for the hottest day of the year. Twenty of the 30 rooftop units had 2 (hermetically sealed) compressors, typically 5 or 7.5 ton each. In their related zones, two-stage T-stats would bring on the first and then the second compressor as needed. On morning startup or on very hot days, however, the T-stats would bring the second compressor on, even though the general cooling overcapacity in the building would have eventually (in about 15 min) taken care of the entire load without the second compressor's being needed. That is, the two-stage T-stat was unduly "impatient."

The overcapacity was corrected as follows:

- To reduce the excess capacity, the second-stage low-voltage control wire was simply cut, and the open ends were taped off. The cut was made about 1 in. back from the screw terminal connector, leaving an "audit trail" for future service personnel in case anyone ever wanted to reinstate the redundant disconnected compressor. Indeed, the "extra" compressor then became a potential spare unit that could easily replace the other compressor should it ever fail.
- It can be appreciated that each modified roof top unit would be left running more air across both the condenser coil and the DX coil than was actually needed, but even so, the improvement in occupant comfort and electric demand reduction was dramatic. Fig. 3 shows the two-compressor unit with the cover removed.

Fig. 3 Two-compressor HVAC package unit where one compressor was not needed.

Resolve Systems Fighting Each Other

Days in the spring and fall present probably the most challenging building-comfort control issue. Building managers sometimes cope by running both heating and cooling systems at the same time, sometimes for weeks. This situation almost always presents significant energy saving opportunities, which may be hidden if the heating and cooling systems fight each other to a comfortable standoff.

Approaches to break the standoff include automating the summer–winter cutover. Relatively inexpensive boiler controls use inputs from outside air temperature and/or the space ambient temperature to decide when to go into so-called warm-weather shutdown. Setting the chiller's startup temperature a few degrees higher than the boiler's warm-weather shutdown point will create a useful dead band (i.e., a gap in temperature between shutting down the boiler and starting the chiller).

The dead-band concept leads to significant energy savings, because neither active heating nor active cooling is going on.

Relocate Thermostats to the Areas That They Serve

A company renovated its space by building out new offices but leaving the dropped ceiling unchanged. After the renovation, HVAC diffusers in the ceiling served some areas totally walled off from their controlling thermostats. As a result, certain spaces became grossly overcooled, with their ambient temperature dropping as low as 62°F (in the summer) because there was no way to close the feedback loop from the T-stat to the HVAC unit.

A project to relocate T-stats to their controlled spaces was completed for less than $500. The payback was estimated at about 2 months, although the real payback was improved employee comfort.

Reduce Lag/Lead Triggered by Thermostat Cases

To prevent tampering, vented steel boxes (which are commonly sold for this purpose) had been installed over thermostats in a company's office space. Even so, there were occupant complaints of being too hot and too cold.

By placing small data loggers inside and outside the metal boxes, and then comparing their readings on the same chart, it was discovered that the space ambient temperature was oscillating wildly around the set point. The following scenario was identified:

1. On summer-morning startup, the HVAC unit would cool the space to the comfort-level set point of 73°F in 20 min.
2. Because the metal box cover was initially warm and had considerable thermal mass, however, it took about 40 min for the space temperature signal to reach 73°F.

3. This introduced a 20-min lag in the feedback loop, thus allowing the HVAC to overshoot grossly. The space temperature would drop to 62°F by the time the metal cover and the enclosed thermostat were down to 73°F.
4. At this point, the HVAC would cycle off, but the process would begin again in the opposite direction as the room heated. The metal box kept the thermostat cool until the room was nearly 80°F.

This energy-saving and comfort-improvement opportunity was addressed simply by removing the metal boxes and using stop lugs (which were found already available inside the thermostat) to limit manual tampering to plus or minus 2°F. Then the T-stat, with its quick time constant thus restored, could control the room's ambient temperature accurately.

Use Small Dedicated AC Units to Cool Server Closets, Rather Than Run the Whole Chiller Plant

A company scheduled an HVAC operation in its 90,000-sq.-ft. building during normal business hours. The HVAC system included a 300-ton chilled-water plant supplying some 14 air handlers.

After the building was wired for local area network (LAN) access, two network server closets (small rooms, actually) were provided for LAN infrastructure equipment; these closets required air conditioning continuously. At that point, the company began to run the two air handlers serving the equipment areas on a 24/7 basis. This also meant, of course that the 300-ton chiller plant and associated pumps would be on continuously. Moreover, because only 2 of the 14 air handlers were running, the 300-ton chiller plant tended to short cycle, even after the compressor controls had unloaded as many cylinders as they could.

To address the energy saving opportunity of shutting off the chiller plant and its associated large pumps, a dedicated 2.5-ton AC unit was installed in each server room. These were "split mini" units (air handlers wall mounted and cooled by a DX coil).

Fig. 4 Dedicated AC air-handler DX unit in computer server room.

Total installed cost (the two units themselves, plus installation labor for both) was less than $12,000. Because the units removed the need to cool two entire building zones during weekends and off hours, however, the payback period was estimated at less than 6 months. Fig. 4 shows a typical wall-mounted air handler, the DX side, as installed.

Tighten Leaky Outside Air Dampers

Logger data from morning warmups taken in each of 7 zones (relating to the 7 air handlers in a company's building) showed that 5 of the zones reached comfortable levels in 40 min, whereas 2 of the zones took nearly twice as long. When the warmup curves were plotted on the same graph, distinctly different slopes were shown. It was appreciated that the original capacities of the air handlers might be different for each area, of course, but CO_2 monitoring also showed much lower levels of CO_2 concentrations in the slow-warming spaces even though occupancy density was about the same.

These findings suggested that the outside air dampers in the two long-warmup zones were not closing or were open way above the normal minimum for fresh air.

Subsequent inspection showed that in one of the suspect zones, the damper actuator linkage was slack, and in the other zone, the actuator shaft had broken off inside the air handler.

Repairs were made in less than 2 h, and the heating bill dropped by about 8%, giving the project a payback period of less than 1 week.

Reduce Fan Speeds by Adjusting Variable-Speed Pulleys When Appropriate

Reducing fan speeds (consistent with ASHRAE 62[2] and building comfort requirements) is a common energy saving opportunity. Although it can be done, when applicable, by retrofitting with variable-speed drives, a lower-cost approach is simply to adjust the variable-speed pulley that often already exists on the drive motor shaft.

Pulley adjustment produces a fixed reduction, as opposed to the wide range of variable speeds provided by new variable-speed drives, but the savings impact vs the difference in cost between the two approaches often suggests the low-cost pulley adjustment as at least a first step. Then the business case can be evaluated to see whether there are enough savings left to justify new drives.

Reset on Ambient Air Temperature When Appropriate

A company's HVAC VAV system was designed to provide a single fan discharge temperature of 60°F and

Fig. 5 Use of T-fitting to pick off fan discharge temperature signal to control fan coil valves.

then rely on the VAV box in each zone to control the space by releasing more or less of this fixed-temperature air (with a preset maximum and minimum amount). This is very near to running the system "open loop" except for the (considerable) action of the VAV boxes.

There were days and conditions during occupied periods, however, when no zone in the entire building needed 60°F air—when all zones became cold on a winter morning, for example. (This discussion assumes that the building had already undergone a warmup cycle where all VAV boxes and fan coil hot water valves were 100% open.)

An energy-saving and comfort-improvement solution is to control fan discharge air temperature based on ambient air temperature. Under this approach, there are two complementary feedback loops in play; the return air T-stat informs fan discharge air temperature, and the zone T-stats control the VAV boxes. The savings result from the fact that the VAV boxes will, on average, not need to try so hard to control the space, because they will have more appropriate air. Thus, they will be more closed (i.e., expelling less air and, thus, reducing fan horsepower). This assumes that fan volume is reduced by throttling or by lowering fan speed.

Fig. 5 shows how the pneumatic signal for Fan Discharge Temperature was picked off simply by tapping into the line to the Fan Discharge Temperature panel gauge line. The signal was routed to the existing Receiver-Controller.

Optimize the Strategy for Manual Light Switches

Short of installing occupancy sensors, companies have used a number of low-cost approaches that can save energy on lighting. They include

- Relocate light switches to the spaces they serve. Light switches sometimes are located far from the spaces they control—in some cases, on other floors. Understandably, this leads users to turn lights on, just to be sure, even when they will not be working near the lights. Consider relocating switches to obvious positions close to where users will enter and leave the affected space.
- Discourage users from turning on every light in the place when the first one arrives. When switches are arranged in banks of 4, 6, or even 8, there is a strong tendency of the first person who arrives each morning to turn all the switches on, even lighting a whole floor for an extra 1 or 2 h per day. By labeling the switches with the zones they cover (and/or relocating them as just described) and by suggesting to users that they turn on just what they need, important savings can be found.
- Leave a few selected fluorescent units on 24/7 as night lights. Understandably, users leaving a lighted space at night would not turn off lights if that left them in total darkness. An energy-saving solution is to provide lighting levels sufficient for a safe exit by identifying intermittently spaced fluorescent (or other approved energy-efficient) lights to be wired on 24/7. Although some energy will be used all the time, on balance, there can be savings from the larger majority of lights that can then be turned off. When adequate night lights are implemented, place friendly signage by key light switches reminding the last person out to shut off the lights.

Provide Actionable Cost-Accounting Feedback to Users

When the company's cost-accounting system inadvertently hides energy costs from users and managers, savings opportunities will invariably be missed.

A cost-accounting system with the following features can go a long way toward addressing this situation:

- Show both dollars and usage (gals, therms, kWh, kW, etc.) on the same report.
- Separate the impact of varying prices from the impact of varying usage.
- In a column adjacent to actual dollars and usage, display dollars and usage adjusted for weather, number of days in the billing period, and/or changes in the space served. The impact of weather can be estimated by setting up a regression model in which at least one of the independent variables is heating degree days or cooling degree days, thus allowing an apples-to-apples comparison over different periods.
- Track the overall building impact of energy-saving improvements by comparing results to a base year (e.g., the year before a formal energy initiative began).

Use a realistic but easy-to-understand measure, such as Btu per square foot.

CONCLUSION

This paper explored the consequences of implementing low-cost/no-cost energy-saving projects before considering longer-payback projects. A number of examples of such low-cost/no-cost projects and approaches were reviewed.

When a company lets energy projects compete for funds in terms of the whole company's needs, as opposed to considering only the impact on the energy or building services department, low-cost/no-cost approaches, done first, create a reordering of project priorities. Indeed, the success of low-cost approaches may even make subsequent investment in long-payback projects unwarranted.

REFERENCES

1. ASHRAE Guideline 12-2000. *Minimizing the Risk of Legionellosis Associated with Building Water Systems*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.: Atlanta, GA.
2. ASHRAE Standard 62.2-2004 *Ventilation for Acceptable Indoor Air Quality*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.: Atlanta, GA.

Energy Efficiency: Strategic Facility Guidelines☆

Steve Doty
Colorado Springs Utilities, Colorado Springs, Colorado, U.S.A.

Abstract
The intent of this article is to offer a convenient list of strategic guidelines to help steer new building designs toward energy efficiency. It is hoped that the project owner, interested in achieving greater than average savings, will provide this list to the design team as part of their project intent instructions. To encourage its use and acceptance, emphasis is placed on keeping the list at a manageable size, with only proven solutions.

Through a top-down approach, using this document as a convenient tool, building owners can set the expectations of energy frugality, create positive changes in building practices, and reward themselves with operational savings.

INTRODUCTION

The old saying "an ounce of prevention is worth a pound of cure" is very applicable to creating energy efficient buildings. The economic realities of retrofitting existing buildings usually result in only the worst ones getting attention, with the average ones left to their life cycle. By contrast, evaluating improvements to designs that exist only on paper has a much smaller economic barrier to change, that being the differential cost from the standard designs. This simple fact makes it easier to create a new efficient building than forcing an older one to be more efficient.

A one-size-fits-all approach is not expected to work for every area of the country, nor for every business segment. Design practices, standard of care, and things that make practical sense will vary by region along with the climate differences and the business it serves. The content of this document was originally prepared for building owners in the Colorado Springs area, and will need local amending for use in other areas like Miami, Florida (all cooling, no heating), and Alaska (all heating, no cooling), and especially in areas with persistent high humidity. Still, many of the concepts presented will be applicable to many areas. One referenced text[1] provides recommendations by climate zone, acknowledging the regional differences. Local adjustments aligned with the eight climate zones are suggested, but beyond the scope of this text. As a starting point, basic allowances have been made to adapt climate-dependent criteria to other areas as follows:

- For most humidity-dependent heating ventilating and air conditioning (HVAC) operations, a distinction is made for climates that are either below 65°F or above 75°F wet bulb temperatures. Interpolation and judgment for points in between is required.
- Evaporative cooling constraints are unique and presented very conservatively, erring on the side of occupant comfort. The climate distinction presented is outdoor air consistently below 52°F wet bulb and below 42°F dew point. (This will no doubt be debated by some designers in semi-arid climates. However, experience has shown that at moisture levels much above these values it becomes increasingly hard to *guarantee* comfort all the time. Since the choice to use evaporative cooling is a large fork in the road for the customer, presenting conservative parameters is deliberate and are sure to work. If evaporative cooling is chosen in "fringe" climates, a supplemental conventional cooling system integrated into the first stage of evaporative cooling is suggested).
- Glazing U-value limits are relaxed in areas where HVAC heating is not used. Note that this suggests double pane windows anywhere HVAC heating is used.
- Glazing shading coefficients are relaxed in areas where HVAC cooling is not used, but are otherwise strict. Solar loads drive peak electrical demands if nothing else.

These climate-dependent items will no doubt be a work in progress, and so they are marked with an asterisk (*) with suggested criteria in *italics* for convenience in editing. I did my best!

To the owner who may use this: not all of the suggestions will apply to all buildings, but many of them will. Don't worry that a number of these won't make sense in your particular building—but do ask the design team to

☆ This entry originally appeared as "Strategic Facility Guidelines for Improved Energy Efficiency in New Buildings" in *Strategic Planning for Energy and the Environment*. Reprinted with permission from AEE/Fairmont Press.

Keywords: Commissioning; Strategic facility guidelines; Sustainability; Internal rate of return; Simple payback; Energy transport; Energy budget; Maintenance activities; Optimization; Evaporative cooling; Comfort envelope; Integrated design; Guide specification; Downsizing; Test and balance; Thermal breaks; Simultaneous heating and cooling.

explain which measures do not apply, and why. Doing so will reinforce your intentions to achieve operational savings by smart design decisions up front, and motivate the design team to make a good effort to accommodate.

THE FACILITY GUIDE SPECIFICATION AND THE TOP-DOWN APPROACH

The approach used for this article is to tabulate a list of suggested "Dos and Don'ts," suitable for use in a facility **guide specification**. A guide specification is a hand-out document given to a design team at the beginning of a project to provide general instructions and owner preferences. The owner handing out the guidelines has more of an effect on the end result than the same suggestions made from a lone team member, hence the term **top-down**. These instructions are then integrated into the other governing documents and codes that eventually form the design. Traditional guide specifications, used by national accounts, campuses, and large facilities, spell out preferred manufacturers, acceptable types of piping, valves, light fixtures, pavement, etc. The concept of energy efficient guidelines in the owner's guide specification is a natural and overdue extension of an existing document. Even if the owner does not have a guide specification to add this to, the listed items in this document can be used in stand-alone fashion to serve the same purpose and provide the same benefit to the owner.

More Top-Down Features

Beginning at the point of use is a common theme for all energy engineers. In the case of the office building, stipulating temperature values can have a pronounced effect on energy use. The indoor temperatures become design parameters that calculations and equipment sizing hinge upon. A design that requires 72°F summer indoor conditions will be larger and use more energy than one designed for 76°F, and yet with proper attire and humidity control, 76°F is a very reasonable temperature. A 72°F design can save energy by turning up the thermostat, but a 76°F design will additionally conserve first-cost dollars. Many designers are reluctant to push the comfort envelope for fear of reprisal against them, and that is where the top-down support concept comes in. With management support, and some education and encouragement, paradigms can slowly change.

Integrated Design

With more creative design solutions such as **integrated design**, more options are apparent to design teams now than in the past. This concept acknowledges cases where decisions in one trade (e.g., electrical or architectural) benefit another trade (e.g., mechanical). It may seem odd, but traditional design practices are compartmentalized by trade, and haven't always allowed the big-picture benefits to find their way to the owner. This is partly due to traditional fee structures where design fees are influenced by the cost of construction; the resistance to downsizing anything becomes obvious in this light. The integrated design process allows an extra expense in one trade to be considered if it produces a corresponding savings in another trade. To avoid redesign costs, this is best implemented with emphasis on group schematic designs (to air out the ideas early), as well as requiring good cost-estimating skills on the design team. A good architect is needed who can be receptive to new ideas, flexible, encourage the integration process, and also know enough to call a halt to the options at some point to maintain a reasonable schedule and finish the job without escalating design fees. By carefully selecting the design team members, constructive opportunities will come from this process which are usually well worth the effort. Examples of integrated design concepts follow, many of them revenue-neutral with sustained energy benefits for "free."

- The added cost of high performance suspended film windows can be offset in some climates by the elimination of perimeter fin-tube heating.
- Downsizing cooling systems in conjunction with reduced lighting power design.
- Downsizing cooling systems in conjunction with improved window shading coefficients, films, or exterior shading systems.
- Downsizing heating and cooling systems by using 1% or 2% ASHRAE design outdoor weather conditions[2] instead of 0.4%, allowing the temperature to drift up a few degrees for a few hours of the year.
- Downsizing a boiler or hot water unit by virtue of selecting higher efficiency equipment. Since output is the design driver, it is often possible to utilize the next smaller size unit, but at higher efficiency, to achieve the same result.
- Downsizing primary heating and cooling equipment in conjunction with upgrades in envelope elements like insulation or, especially, window shading.
- Downsizing primary heating and cooling equipment in conjunction with heat recovery systems.
- Downsizing fan and pump motors, and primary cooling equipment, by increasing duct and pipe sizes, filter areas, coil areas, etc. as the trade-off for using less transport energy. Note that the extra heat of transport energy elements, especially fans, often drives the equipment size up a notch.
- Downsizing overhead lighting and HVAC cooling via an owner commitment to a greater use of task lighting.

Table 1 Equivalent rate of return values for various simple payback

	ECM PROJECT LIFE			
	5YEAR	10 YEAR	15 YEAR	20 YEAR
SIMPLE PAYBACK	IRR	IRR	IRR	IRR
2	40.0%	49.0%	50.0%	50.0%
3	**20.0%**	31.0%	33.0%	33.0%
4	8.0%	**21.0%**	23.5%	24.6%
5		15.0%	**18.0%**	19.4%
6		10.6%	14.5%	**15.8%**
7		7.0%	11.5%	13.0%
8		4.3%	9.0%	11.0%
SPB typ 9		1.8%	7.1%	9.2%
10-yr max 10			5.5%	7.7%

Simple Payback vs. Internal Rate of Return
payback periods noted for various project lifespans that achieve greater than 15% rate of return

Derivation: Internal Rate of Return (IRR) is that interest rate where the present worth of the savings is equal to the initial investment.

$P = A * (P/A, i, n)$

$$P = A * \frac{(1+i)^n - 1}{i\,(1+i)^n}$$

so, for some value of i,

$$P/A \text{ (simple payback)} = \frac{(1+i)^n - 1}{i\,(1+i)^n}$$

In addition to downsizing, integrated design can be used as an investment tool, with utility use reduction as the return. If sufficient funding exists, the owner can allow upgrades with identifiable costs and annual savings to be proposed, with a stipulated payback period such as 2–5 yr, to capture the fruit that is just above the proverbial low-hanging level. For example, an upgrade with a 10 yr life and a 4 yr simple payback is an equivalent internal rate of return of 21%, which is an attractive return for investment dollars.

To the energy engineer: be *sure* of your calculations, and even derate them, before leading your owner out with borrowed money (Table 1).

Sustainability

The guideline includes instructions to the owner's staff as well as designers and contractors. This is because the construction of a well-performing building is not the end of the story. The energy and water use of the building will also depend upon user habits at the points of use, and interaction with the building occupants is a vital component for program success. While many endeavors begin a long slide in performance after inception, it is a further intent of this document to encourage **sustainability** of operations. To this end, the suggestions for thorough documentation, commissioning, maintenance, and occupant education are included, and require an ongoing commitment by the owner.

STRATEGIC FACILITY GUIDELINES FOR IMPROVED ENERGY EFFICIENCY IN NEW BUILDINGS

Purpose

- Implementation of the contents of this guideline will reduce facility energy use by 30–50% when compared with ASHRAE 90.1 Base Building and Minimum Local Energy Codes.
- Suggested use of this document is for the project owner interested in achieving the stated savings to provide to the design team as part of their project intent guidance.

General

- Design document submittals must include detailed narrative descriptions of system functionality, features, limitations, design assumptions, and parameters, for use by the owner. The narratives will be detailed enough to provide benefit to subsequent design teams, and will be written to be informative and useful to building operations personnel. The narrative will be provided as a deliverable with the schematic design, and will be updated with each subsequent design delivery including design development (DD) and construction documents (CD) phases. In its final form, this document shall be placed on the first sheet of the drawing set behind the title page, so that the information is retrievable years later when all that is available to facility operations are the drawings. Design assumptions include number of people, indoor and outdoor HVAC design conditions, foot-candles of illumination, hours of operation, provisions for future expansion (if any), roof snow load, rainfall rates, etc. that all define the capabilities of the building.
- All equipment schedules, including HVAC, plumbing, lighting, glazing, and insulation shall be put onto the drawings and shall not reside in the specification books, so that the information is retrievable years later when all that is available to facility operations are the drawings.
- Design thermal insulation values and glazing properties that affect energy use (U-value, shading coefficient, etc.) shall be clearly noted on the drawings.

- Project commissioning that includes identifying measurable energy savings goals and monitoring the design and construction activities with these as project intent items, with early detection and notification of any project changes that impact energy use or demand.
- Project final payment contingent upon:
  - Receipt of accepted accurate as-built drawings, with accuracy verified by owner and signed by the contractor.
  - Receipt of accurate and complete operations and maintenance (O/M) manuals, with certified factory performance data, repair parts data, and vendor contact information for all energy consuming equipment, including all HVAC and lighting equipment and controls.
  - Receipt of test and balance report that demonstrates design intent is met for air, water, and ventilation quantities, showing design quantities, final adjusted quantities, and percent variance. This would include all variable air volume (VAV) box minimum settings shown, including both heating and cooling balanced air quantities. This would also include any equipment performance testing that was specified for the project.
  - Verification by the owner that the test and balance settings include permanent markings and hence these settings can be preserved over time.
  - Receipt of on-site factory-authorized start-up testing for primary HVAC equipment including chillers and boilers, with efficiency and heat/cool performance figures and heat exchanger approach temperatures to serve as baseline. The submitted reports would include as a minimum heating/cooling output, gas/electric energy input, heat exchanger approach temperatures, water and air flows.
  - Receipt of control shop drawings with detailed descriptions of operation.
  - Acceptance testing of the automatic control system using the approved sequence of operation, and verification that the sequences are fully descriptive and accurate. Acceptance testing also includes review of the control system man-machine interface provisions to become familiar with each adjustable point in the system. Acceptance is by the owner, who will witness each sequence as part of the turnover training requirements.
- Building design must prevent negative pressure condition, unless safety considerations require it.
- Electric resistance space heating, air heating, and water heating are not allowed, unless there is no means to get natural gas to the site.
- Portable space heaters not allowed, unless required for an approved emergency measure.

Energy Use, Overall Performance

- Using ASHRAE 90.1 or local energy code as a baseline, demonstrate through computer modeling that the building energy use will be at least 30% less than this value.

Irrigation Water Use, Overall Performance

- Using standard Kentucky Bluegrass sod and average regional rainfall rates as a baseline, demonstrate that irrigation use for the property will be 50% or less of this value.

Test and Balance

- Balance using "proportional balancing," a technique that strives to reduce throttling losses, which permanently energy transport penalties (pump and fan power).
- Any motor over 5 hp found to be throttled with a resistance element (valve or damper) more than 25% must be altered by sheave change or impeller trim to eliminate lifelong energy waste from excessive throttling losses.
- All 3-phase motor loads, including HVAC equipment, must include voltage balance verification as part of the test and balance (TAB) work. Voltage imbalance of more than 1% indicates unbalanced electrical service in the building and unacceptable efficiency losses.
- Vertical return air shafts serving multiple floors require a balancing damper at each branch outlet to proportion the return air by floor.
- Air flow performance testing for all Air Conditioning and Refrigeration Institute (ARI) certified HVAC factory packaged unitary equipment greater than 5-tons capacity. Heating and cooling performance and efficiency verification is assumed via the ARI certification process.
- Heating efficiency, cooling efficiency, and air flow performance testing for all HVAC *split system* equipment greater than 5-tons capacity or 200,000 Btuh input heating capacity.
- Water flow performance testing for all ARI certified factory packaged water chillers. Cooling performance and efficiency verification is assumed via the ARI certification process.
- Water flow and combustion efficiency testing for all boiler equipment.
- Combustion efficiency testing for all boiler equipment unless factory startup is provided on site.
- Cooling tower thermal performance verification is assumed via the Cooling Tower Institute (CTI) certification process.

Electrical Service

- Provide separate utility metering for electric, gas, and water for the building, separate from other buildings.
- Electrical transformer conversion efficiency not less than 95% efficient at all loads from 25% to 100% capacity. Dry-type transformers National Electrical Manufacturers Association (NEMA) TP-1 compliant.
- Locate transformers in perimeter areas that do not require air conditioning for cooling.
- Power factor (PF) correction on large motor loads, for overall building PF of 90% or better. Large mechanical equipment can be provided with the correction equipment. If motor loads are segregated, this can be done at the switchgear.
- Arrange switchgear and distribution to allow metering of the following electrical loads (requires segregating loads):
  — Lighting.
  — Motors and Mechanical.
  — Plug Loads and Other.

Envelope

- Orient buildings long dimensions E–W where possible to reduce E–W exposure and associated solar load.
- Provide building entrance vestibule large enough to close one door before the next one opens (air lock).
- Where thermal breaks are used, the thermal break material must have thermal conductivity properties an order of magnitude better than the higher conductivity material it touches, and must be at least 1/2 in. thick.
- *Minimum wall insulation 25% *beyond ASHRAE 90.1 values, but not less than R-19*. Insulation is generally not expensive during new construction. Incorporate exterior insulation system (outboard of the studs) for at least one-half of the total R-value, to avoid thermal short circuits of standard metal stud walls, which derate simple batt insulation system by approximately 50%, e.g., a standard stud wall with R-19 batts between the studs yields an overall R-9.5.
- *Minimum roof insulation R-value 25% *beyond ASHRAE 90.1 values, but not less than R-30*. Insulation is generally not expensive during new construction. Select insulation that will retain its thermal properties if wet, e.g., closed cell material.
- Glazing meeting the following requirements:
  — Thermal breaks required.
  — *U-factor of 0.35 or less *where HVAC heating is provided*.
  — *Low-E coatings on east- and west-facing glass *where HVAC cooling is provided*.
  — *Max shading coefficient of 0.2 *where HVAC cooling is provided*. Note: any combination of tinting, coating, awnings, or other exterior shading can be used to achieve this. This is to say that no more than 20% of the heat energy from the sunlit glazing is to get into the building.
  — Glazing not more than 25% of gross wall area.
- Skylight/Clerestory elements must meet the following requirements:
  — Thermal breaks required.
  — Triple pane (layer) construction with sealed air space(s).
  — Overall U-value of 0.25 or less.
  — *Skylight shading coefficient must be 0.2 or less *where HVAC cooling is provided*.
  — *Low-E coating *where HVAC cooling is provided*.
- Skylight/Clerestory area not to exceed 5% of roof area.
- Return plenums and shafts designed with an air barrier for leakage not exceeding 0.25 CFM/ft^2 of building envelope surface area at 50 Pa (Energy Efficient Building Association (EBBA) criteria). Shaft construction requires field testing and verification.
- Building envelope devoid of thermal short circuits. Provide thermal break at all structural members between outside and inside surfaces.
- Building leakage testing required (new buildings), with no more than 0.25 CFM/ft^2 of building envelope surface area at 50 Pa (EBBA criteria).
- Utilize lower ceilings to reduce necessary light input power for equivalent light levels at the work surface.
- Utilize reflective (light) color interior colors for ceilings, walls, furniture, and floors, to allow reduced lighting power for comparable illumination. It can take up to 40% more light to illuminate a dark room than a light room with a direct lighting system.
- Good reflectance parameters to use when picking interior surfaces and colors follow. If these values are used and the lighting designer is informed of it, the integrated design process will allow reduced lighting power to achieve the desired light levels.
  — Min 80% reflective ceiling.
  — Min 50% reflective walls.
  — Min 25% reflective floor and furniture.
- Provide operable blinds for vision glass.

Lighting

- Follow ASHRAE 90.1 or local energy code requirements for lighting power budget guidelines, and verify that designs are lower than these limits while meeting current applicable Illuminating Engineers Society of North America (IESNA) lighting illumination requirements.
- Utilize task lighting and less on overhead lighting for desk work.
- Provide separate circuits for perimeter lights within 10 ft of the wall, to allow manual or automatic light harvesting.

- Use 1-2-3 switching for large open interior area spaces.
- Use ballast that will tolerate removing at least one bulb with no detriment.
- Where occupancy sensors are used, provide "switching ballast" that will tolerate large numbers of on–off cycles without bulb or ballast life span detriment.
- Use electronic ballast instead of magnetic ballast.
- Use ballast factor in the lighting design to improve lighting system efficiency. Because the ballast mostly determines how many watts are used, ballast choice is critical to achieving best energy efficiency.
- Coordinating light output with "ballast factor" is an excellent tool for providing optimum light levels and energy use.
- Use high power factor ballast, with minimum PF of 95% at all loads.
- Occupancy sensors in conference rooms, warehouses, and multifunction rooms. Also in locker rooms and restrooms, but with some continuous manual switched lighting in these areas.
- Photo-cell controlled lights in the vicinity of skylights.
- Do not use U-tube fluorescent lights, due to high replacement bulb costs.
- Do not use incandescent lights.
- Outdoor lighting on photocell or time switch.

Motors and Drives

- All motors meet or exceed EPACT-1992 efficiency standards.
- Variable frequency drive (VFD) on all HVAC motors larger than 10 hp that have variable load.
- Motor nameplate horsepower (hp) not more than 20% higher than actual brake horsepower served (i.e., do not grossly oversize motors).

HVAC

- Provide HVAC calculations and demonstrate equipment is not oversized. Equipment selection should not be more than 10% greater capacity than calculated values indicate.
- HVAC calculations will include both maximum and minimum heat/cool loads and equipment shall be designed to accommodate these load swings, maintaining heat/cool efficiency equal to or better than full load efficiency at reduced loads down to 25% of maximum load, e.g., equipment capacity will track load swings and energy efficiency will be maintained at all loads.
- Provide necessary outside air (OA), but no more than this. Excess ventilation represents a large and controllable energy use. Reduce exhaust to minimum levels and utilize variable exhaust when possible instead of continuous exhaust. Reduce "pressurization" air commensurate with building leakage characteristics. If the building is tested to low leakage as indicated herein, there should be little need for this extra air, or the heat/cool energy it requires. Design controls to dynamically vary outside air with occupancy.
- VAV box primary heating cubic feet per minute (CFM) shall be not higher than the cooling minimum CFM. This is to say the VAV box primary damper will *not* open up during heating mode.
- Zoning:
  - Design HVAC zoning to require heating *or* cooling, not both. This will improve comfort and also reduce the inherent need for simultaneous heating and cooling.
  - Do not zone any interior areas together with any exterior areas.
  - Do not zone more than three private offices together.
  - Do not zone more than one exposure (N, S, E, and W) together.
- Design and control settings for ASHRAE Standard 55 comfort envelope, which indicates 90% occupant comfort. Appropriate temperatures will vary depending on humidity levels. For example, in Colorado Springs (dry) the following space temperatures are appropriate:
  - 71°F heating.
  - 76°F cooling.
  - Facilities may institute a range 68–72°F heating and 74–78°F cooling, provided that a 5° dead band is kept between the heating and cooling settings.
- Do not heat warehouses above 60°F.
- Do not cool data centers below 72°F.
- Do not use electric resistance heat.
- Do not use perimeter fin-tube hydronic heating.
- **In cooler climates* where HVAC economizers are used, designs should normally favor air-economizers over water-economizers since the efficiency kW/ton is better for the air system. The water economizer "free cooling" includes the pumping and cooling tower fan horsepower, as well as the air handler fan. If the air handler fan power is considered required regardless of cooling source, the air-side economizer is truly "free" cooling.
- **In very dry climates, with outdoor air wet bulb temperatures consistently less than 52°F and dew point consistently less than 42°F*, evaporative cooling (direct, indirect, or direct-indirect) should be used in lieu of mechanical refrigeration cooling, as long as indoor humidity of 40% rH or less can be maintained. To the water consumption issue, it is this author's opinion that water is a renewable resource and does not disappear from the planet like fossil fuels do, and hence this technology should be used without environmental resources concern.
- Packaged HVAC cooling equipment not less than SEER-13 or EER-12, as applicable.

- *Air-side economizers for all rooftop equipment, regardless of size, *for climates with design wet bulb temperatures below 65°F.*
- Avoid duct liner and fiber-board ducts due to higher air friction and energy transport penalties.
- Insulate all outdoor ductwork to R-15 minimum.
- Use angled filters in lieu of flat filters, to reduce air friction loss.
- Reduce coil and filter velocities to a maximum of 400 fpm to lower permanent air system losses and fan power.
- Avoid series-fan-powered VAV boxes.
- For fan-powered VAV boxes, use energy conservation measure (ECM) motors to achieve minimum 80% efficiency. Although the motors are not large, when there are many of them this efficiency benefit is significant.
- Heat recovery for any 100% outside air intake point that is greater than 5000 CFM when the air is heated or cooled.
- Air filter requirements:
  - Terminal units (fan coils, fan powered boxes, unit vents): 20% (1-in. pleated). Note: this may require an oversize fan on small terminal equipment, and not all manufacturers can accommodate.
  - Air handlers with 25% or less OA: 30%—MERV-7
  - Air handlers with 25–50% OA: 45%—MERV-9
  - Air handlers with more than 50% OA: 85%—MERV-13
  - Provide manometers across filter banks for all air handlers over 20 tons capacity. Equip manometers with means to mark the "new-clean" filter condition, and change-out points.
- *Air-cooled condensing units over 25 tons, provide evaporative precooling, *for climates with design wet bulb temperatures below 65°F.*
- Make-up meter for all hydronic systems to log system leaks and maintain glycol mix.
- Separate systems for 24-7 loads to prevent running the whole building to serve a small load.
- *Direct evaporative post cooling for all chilled water systems, *for climates with design wet bulb temperatures below 65°F.*
- Require duct leakage testing for all ducts 2 in. w.c. design pressure class or greater.
- For process exhaust and fume hoods, design for variable exhaust and make-up.
- Utilize general exhaust air as make-up for toilet exhaust and other exhaust where possible.
- Dedicated outside air system (DOAS) for large office facilities (over 50,000 SF) with VAV systems, allowing ZERO minimum settings for all VAV boxes. This will eliminate the VAV reheat penalty, and the internal zone over-cooling effect from VAV minimums which often requires running the boilers throughout the year for comfort control.
- Separate interior and exterior VAV zoning for open-plan rooms to utilize zero-minimums in the interior spaces.
- Do not use grooved pipe fittings in hydronic heating or cooling piping systems to prevent operating central heating and cooling equipment year-round on account of these fittings.
- Verify that all manufacturer's recommended clearances are observed for air cooled equipment.
- Humidification:
  - Do not humidify any general occupancy buildings such as offices, warehouses, or service centers.
  - In data centers *only*, humidification should not exceed 30–35% rH.
  - Where humidification is used, humidifiers should be ultrasonic, mister, or pad type, and should not be electric resistance or infrared type.
  - Do not locate humidifiers upstream of cooling coils, to avoid simultaneous humidification—dehumidification.
  - Where humidification is used, provide for elevated apparatus dew point of cooling coils or other means to prevent simultaneous humidification—dehumidification.
- Dehumidification:
  - Do not dehumidify below 45% rH.
- Provide performance and efficiency testing of package heating and cooling equipment over 7000 CFM or 20 tons or 500,000 Btu input heating units with factory authorized equipment representatives. Test figures to include on-site gross heat/cool output, fuel and electrical input, and efficiency, compared to advertised values.

Energy Transport Systems—Energy Budget

- For HVAC air systems, the maximum energy transport budget will be:
  - No less than 10 Btu cooling and heating delivered to the space per Btu of fan energy spent at design conditions.

    This will generally steer the design toward generous sizing of sheet metal ducts, air handler cabinetry, coils and filters, higher efficiency fans (0.7 or better), and higher system differential temperatures to reduce air flow rates, but it will result in greatly reduced lifetime energy use since it lowers the bar of system pressure.
  - Fan hp limitation from:

    Cooling fan hp max input

$$= \frac{\text{Cooling Btu gross output}}{(10 \times 3413 \times \text{motor eff})}.$$

Energy Eff–Serv

— Air hp limitation from:

$$\text{Cooling fan hp max budget} \times \text{fan eff}.$$

— TSP limitation from:

$$\text{TSP} = (\text{air hp} \times \text{fan eff} \times 6360)/\text{CFM}.$$

For example, a 100-ton HVAC air system using 80% e motor, 70% e fan, and 350 CFM per ton would be limited to 44 hp motor load and 3.9 in. w.c. TSP. NOTE: for systems with both supply and return fans, the transport energy considers both combined as the "fan."

- For HVAC water systems, the maximum energy transport budget will be:
  — No less than 50 Btu cooling and heating delivered to the space per Btu of pump energy spent, at design conditions.
  This will generally steer the design toward generous sizing of piping, strainers, coils, and heat exchangers, higher efficiency pumps (0.75 or better) and higher system differential temperatures to reduce water flow rates, but will result in reduced lifetime energy use since it lowers the bar of system pressure.
  — *Pump hp limitation from*:

$$\text{Cooling pump hp max input} = \frac{\text{Cooling Btu gross output}}{(50 \times 3413 \times \text{motor eff})}.$$

— *Water hp budget from*:

$$\text{Pump max hp} \times \text{pump eff}$$

— *HEAD limitation from*:

$$\text{HEAD} = (\text{water hp} \times \text{pump eff} \times 3960)/\text{GPM}.$$

Hydronic Circulating Systems

- Heating: minimum 40° dT design, to reduce circulating flow rates and pump hp.
- Cooling: minimum 16° dT design, to reduce circulating flow rates and pump hp.

Boilers and Furnaces

- No atmospheric burners.
- No standing pilots.
- Design hydronic system coils to return water to the boiler at or below 140°F water with a minimum of 40°F temperature drop. This will reduce circulating pump energy and improve boiler efficiencies.
- Minimum efficiency of 85% at all loads down to 25% load.
- For heating load turn-down greater than 4:1, provide modular boilers or a jockey boiler.
- For multiple boilers sharing multiple pumps, provide motorized valves to cause water flow to occur *only* through the operating boiler.
- Provide stack dampers interlocked to burner fuel valve operation.

Chillers

- *Water-cooled centrifugal efficiency 0.5 kW/ton or less with 70°F condenser water and 45°F chilled water *for climates with design wet bulb 65°F and lower. 0.58 kW/ton or less with 85°F condenser water and 45°F chilled water in climates where design wet bulb temperatures are above 75°F.*
- *Water-cooled centrifugal units able to accept 55°F condenser water at 3 gpm per ton, all loads. *Beneficial in dry climates with design wet bulb temperatures less than 65°F and typical wet bulb temperatures less than 50°F.*
- *Water-cooled positive displacement units 0.7 kW/ton or less with 70°F entering condenser and 45°F chilled water *for climates with design wet bulb 65°F and lower 0.81 kW/ton or less with 85°F condenser water and 45°F chilled water in climates where design wet bulb temperatures are above 75°F.*
- Do not provide chilled water temperatures less than 45°F. Select cooling coils to provide necessary cooling with 45°F chilled water or higher.
- Air-cooled chiller efficiency 1.0 kW/ton or less with 95°F entering air.
- *Air-cooled chillers over 25 tons, provide evaporative precooling *where design wet bulb temperatures are less than 65°F.*

Cooling Towers

- Selected for 7°F approach at design wet bulb and 0.05 kW/ton or less fan input power. This will steer the design toward a larger free-breathing cooling tower box with a small fan, minimizing parasitic losses from the cooling tower fan. Cooling tower fan kW/ton should not be more than one-tenth of the chiller it serves.
- *Set condenser water temperature set point not higher than 70°F *for climates with design wet bulb 65°F and lower. For climates with higher wet bulb temperatures,*

design to 7°F above design wet bulb with reset controls to lower the setting whenever conditions permit.
- Water treatment control for minimum seven cycles of concentration to conserve water.
- Specify cooling tower thermal performance to be certified in accordance with CTI STD-201.

Air-Cooled Equipment and Cooling Towers in Enclosures

- Locate to prevent air short-circuiting and associated loss of thermal performance. Rule of thumb is the height of the vertical finned surface projected horizontally. The fan discharge must be at or above the top of the enclosure, the distance to the enclosure walls should be as indicated above, and there should be amply sized inlet air openings in the enclosure walls as low as possible.

Ground Source Heat Pumps

- Coefficient of performance (COP) 4.0 or higher at 40°F entering water.
- EER 17 or higher at 80°F entering water.
- No electric resistance heating.

Controls

- Design OUT all simultaneous heating and cooling through the use of proper zoning, interlocks, and dead bands. This includes all constant volume systems and terminal unit systems. VAV systems inherently have an overlap which should be minimized by water and air reset in heating season, prudent use of minimum VAV box settings, and consideration of systems that separate the outside air from the supply air (SA).
- Programmed start–stop for lighting and HVAC systems with option for temporary user overrides. Use these controls to prevent unnecessary operating hours.
- Lock out air flows for conference rooms and intermittent occupancy rooms by interlocking VAV box to close with occupancy sensors.
- Lock out chiller operation below 50°F, except for data centers or humidity-sensitive areas that cannot use outside air for cooling.
- Lock out boiler operation above 60°F, unless space temperatures cannot be maintained within the specified ranges any other way.
- *All cooling by air economizer below 55°F *for climates with design wet bulb 75°F and lower*.
- Night setback for heating. Suggested temperature for unoccupied time is 60°F.
- No night set-up for cooling—no cooling operation in unoccupied times for general occupancy buildings. If building temperature rise during unoccupied times can cause detriment, then limit off-hours cooling operation to 85°F indoor temperature.
- Reset boiler hot water temperature settings in mild weather.
- *Reset chilled water temperature settings in mild weather, *provided that outdoor air dew point is below indoor dew point levels*. Refrigeration savings generally exceeds increases in pump power.
- Provide appropriate interlock for all exhaust fans to prevent infiltration of outside air from uncontrolled exhaust fans that operate in unoccupied times.
- All analog instruments—temperature, pressure, etc. other than on–off devices—must be calibrated initially (or verified for non-adjustable devices). Merely accepting out-of-the-box performance without verification is not acceptable.
- Two-year guarantee on calibration, with 18-month re-calibration of all analog inputs.
- Air handler control valves with a residual positive seating mechanism for positive closure. Use of travel stops alone for this is not acceptable.
- For terminal units and heating/cooling hydronic water flow rates less than 10 gpm, use characterized ball valves for control valves instead of globe valves or flapper valves, for their inherent improved long-term close-off performance. This will reduce energy use from simultaneous heating and cooling.
- Valve and damper actuator close-off rating at least 150% of max system pressure at that point, but not less than 50 psid (water) and 4 in. w.c. (air).
- Dampers at system air intake and exhaust with leakage rating not more than 10 CFM/ft^2 at 4 in. water column gage when tested in accordance with Air Movement and Control Association (AMCA) Standard 300.
- Water coil control valve wide open pressure drop sizing not to exceed the full flow coil water-side pressure drop.
- Provide main electrical energy and demand metering, and main gas metering. Establish baseline and then trend; log kBtu/SF, kWh/SF-yr, and kW perpetually and generate alarm if energy use exceeds baseline.
- Implement demand-limiting or load leveling strategies to improve load factor and reduce demand charges. Stagger-start any large loads, e.g., morning warm-up or cool-down events. Use VFDs to limit fan, pump, and chiller loads to 90% during peak hours, etc.
- Independent heating and cooling set points for space control.
- Space temperature user adjustment locked out or, if provided, limited to ±2°F.

- 5°F dead band between space heating and cooling set points to prevent inadvertent overlap at zone heat/cool equipment, and from adjacent zones.
- 5°F dead band between air handler heating and cooling (or economizer) set points, e.g., preheat coil cannot share a single, sequenced, set point with the economizer or cooling control.
- Provide separate lighting and HVAC time schedules.
- For chillers (condenser) and hot water boilers, use temperature sensors to log heat exchanger approach values, to prompt predictive maintenance for cleaning fouled heat exchange surfaces. New-equipment approach will be the baseline value, and approach temperature increases of 50% will prompt servicing.
- Interlock heating and cooling equipment in warehouses serving doorway areas to shut off when roll-up doors are open to reduce waste.
- Optimization routines:
  - — Automatically adjust ventilation rates for actual people count.
  - — Optimal start to delay equipment operation as long as possible.
  - — Demand limiting control point that will limit all VFD-driven air handler fans components to a maximum of 90% max output in summer. This will cause system temperatures to drift up slightly during extreme weather, but will reduce electrical demand for this equipment (and the cooling equipment it serves) compared to full output operation, during times when utility demand is highest. Do not oversize equipment capacity to compensate for this requirement.
  - — Optimal static pressure setting based on VAV box demand, not a fixed set point. This is a polling routine.
  - — **For areas with design wet bulb temperatures below 65°F only*, optimal SA reset that will reset the SA temperature set point upward from 55°F to 62°F for VAV systems during heating season, to reduce reheat energy. This can either be from two methods.
    - * *Method 1*. Basic Optimization. When the main air handler fan is below 40% of capacity and OA temperature is below 40°F.
    - * *Method 2*. Fully Optimized. Polling VAV boxes (at least 80% of the boxes served are at minimum air flows).
  - — Do not reset SA temperature from return air. Do not reset SA temperature during cooling season.
  - — Reset condenser water temperature downward when outdoor conditions permit, using the lowest allowable condenser water the chiller can accept.

Plumbing

- Max shower flow 1.5 gpm.
- Max bathtub volume 35 gal.
- Max urinal water flow 0.5 gpf, or waterless.
- Max lavatory water flow 0.5 gpf.
- Metering (self closing) or infrared lavatory faucets.
- Avoid single lever faucets since these encourage complacency for the use of hot water.
- All domestic hot water piping insulated.
- Heat trap in domestic hot water main outlet piping.
- If a circulating system is used, provide aquastat or timer to prevent continuous operation.
- Max domestic hot water temperature for hand washing 125°F.
- Gas water heaters in lieu of electric where natural gas is available.
- Domestic water heater equipment separate from the building boiler and heating system to prevent year-round operation of central heating equipment.
- Water fountains instead of chilled water coolers.
- Operate the building at reduced pressure (such as 50 psig) instead of 70 psig, to reduce overall usage. Verify that design maintains at least 10 psig over the required minimum pressure at all flush valves.

Management and Maintenance Activities to Sustain Efficiency

- Management support
  - — Create buy-in from the building occupants. Distribute information to building occupants to raise awareness of energy consumption, especially communicating that the user's habits are an essential ingredient to overall success, and are useful and appreciated. This would be in the form of occasional friendly and encouraging reminders of how user participation is helping, fun facts, etc. along with estimated benefits from behavior changes. Provide measured results whenever available.
  - — Enforce temperature setting limitations, including the explanation of why this is helpful and also why it is reasonable. Encourage seasonal dress habits to promote comfort and conservation together.
  - — Prohibit space heaters.
  - — For offices, utilize LCD monitors and the software-driven "monitor power-off" feature, since the monitor represents two-thirds of the whole personal computer (PC) station energy use.
  - — Track monthly energy and water use and maintain annual graphing lines, comparing current and prior years. Establish new benchmark curves after major renovations, alterations, or energy conservation projects. Compare annual use with benchmark and

verify that building energy and water usage per SF is not increasing. Report results to the building occupants as an annual energy use report for their feedback.
— Escrow (save) approximately 5% of the replacement cost per year for the energy consuming equipment in the facility that has a normal life cycle, such as HVAC systems, lighting systems, and control systems. This will allow 20-year replacement work without "surprises" to sustain efficient building operations.
— For leased office space, show the tenants their utility costs to increase awareness and encourage conservation by the users. The typical industry arrangement is to build in utilities into the lease price, so the tenants do not see a separate utility bill. Although the customers are paying for the utilities, having those costs clearly shown will reduce the complacency in utility use.

- Chillers:
  — Owner provides annual equipment "tune up," including cooling efficiency testing and heat exchanger approach measurements.
  — Owner adjusts temperature settings or cleans heat exchangers, or adjusts water flows whenever cooling efficiency tests are less than 90% of new-equipment values. For example, if the new equipment benchmark is 0.5 kW/ton, then a measurement of 0.5/0.905 = 0.55 kW/ton would trigger corrective action.
- Boilers:
  — Owner provides annual equipment "tune up," including combustion efficiency testing and heat exchanger approach measurements.
  — Owner adjusts temperature settings, cleans heat exchangers, or adjusts air-fuel mixture whenever combustion efficiency tests are less than 95% of new-equipment values. For example, if new equipment benchmark is 80%, then a measurement of $0.8 \times 0.95 = 0.76$ would trigger corrective action.
- HVAC air coils:
  — Owner changes filters at least quarterly, and verifies there are no air path short circuits allowing air to bypass the filters.
  — Owner cleans HVAC coils whenever there is any sign of visible accumulation or if air pressure drop is found to be excessive.
- HVAC air-cooled condensers:
  — Owner provides location free from debris, leaves, grass, etc. and adequate spacing for free "breathing" and no recirculation.
  — Owner cleans heat exchange surfaces annually.
- Controls:
  — Owner re-evaluates system occupancy several times each year, to reduce unnecessary HVAC and lighting operating hours.
  — Owner re-evaluates control set points each year including space temperature settings, duct pressure settings, SA temperature settings, reset schedules, and heating and cooling equipment lock-out points.
  — Owner re-calibrates control instruments each two years other than on–off devices.
  — Owner cycles all motorized valves and dampers from open to closed annually, and verifies tight closure.
  — Owner cycles all VAV box dampers from open to closed annually and verifies that the control system is responsive, since these often have a short life and can fail without the user knowing it.

CONCLUSIONS

The listed items in this document are intended to supplement traditional facility guide specifications, as a tool to help steer new building designs toward sustained low energy use.

It is easier to design efficiency into a new building than to retrofit an existing one, for practical and monetary reasons. While we continue to search for ways to upgrade existing buildings, we should influence the new buildings as much as possible for the long term benefits. Engineers and architects alone may understand the benefits and opportunities available, but may not be effective at altering the default course of events for new building designs. The most effective way to assure energy savings as a built-in feature is through a top-down approach where owner support conveys efficiency as a design team priority. The energy-efficient design commitment is made more effective through the use of integrated design, and commissioning can be an effective tool to be certain the design intentions are realized through construction. The sustained energy efficiency goal requires an ongoing commitment from the owner, maintenance staff, and building occupants, and includes training, appreciation, and feedback.

REFERENCES

1. Advanced Energy Design Guide for Small Office Buildings, ASHRAE, 2000.
2. ASHRAE Fundamentals Handbook, 2001.

Energy Information Systems☆

Paul J. Allen
Reedy Creek Energy Services, Walt Disney World Co., Lake Buena Vista, Florida, U.S.A.
David C. Green
Green Management Services, Inc., Fort Myers, Florida, U.S.A.

Abstract
Advances in new equipment, new processes, and new technology are the driving forces in improvements in energy management, energy efficiency, and energy-cost control. Of all the recent developments affecting energy management, the most powerful technology to come into use in the past several years has been information technology (IT). The combination of cheap, high-performance microcomputers and emerging high-capacity communication lines, networks, and the Internet has produced explosive growth in IT and its application throughout our economy. Energy Information Systems (EIS) have been no exception. Information technology and Internet-based systems are the wave of the future. This entry will introduce basic principles, structures, and definitions, and will also show examples for typical EISs.

ENERGY INFORMATION SYSTEMS

The philosophy "If you can measure it, you can manage it" is critical to a sustainable energy management program. Continuous feedback on utility performance is the backbone of an Energy Information Systems (EIS).[1] A basic definition of an EIS is equipment and computer programs that allow users to measure, monitor, and quantify the energy usage of their facilities and to help identify energy conservation opportunities.

Everyone has witnessed the growth and development of the Internet—the largest computer communications network in the world. Using a Web browser, one can access data around the world with a click of a mouse. An EIS should take full advantage of these new tools.

EIS PROCESS

There are two main parts to an EIS: (1) data collection and (2) Web publishing. Fig. 1 shows these two processes in a flow-chart format.

Data Collection

The first task in establishing an EIS is to determine the sources of the energy data. Utility meters monitored by an energy management system or other dedicated utility-monitoring systems are a good source. The metering equipment collects the raw utility data for electric, chilled and hot water, domestic water, natural gas, and compressed air. The utility meters communicate to local data storage devices by preprocessed pulse outputs; by 0–10 V or 4–20 mA analog connections; or by digital, network-based protocols.

Data gathered from all the local data storage devices at a predefined interval (usually on a daily basis) are stored on a server in a relational database (the data warehouse). Examples of relational databases are FoxPro, SQL, and Oracle. A variety of methods are used to retrieve these data:

Modem connection. A modem connection uploads the data from the local data storage device to the energy data server. Typically, the upload takes place on a daily basis, but it may occur more frequently if needed.

LAN or WAN network connection. A local area network (LAN) or a wide area network (WAN) connection established between computers transfers energy data files to the energy data server.

FTP network connection. File transfer protocol (FTP) is an Internet protocol used for transferring files from one computer to another. It moves the energy data files from the local data storage devices to the energy data server.

When the energy data have been transferred to the energy data server, an update program reads all the various data files and reformats them in a format that is used by the Web publishing program.

Web Publishing

To publish the energy data on an Intranet or on the Internet, client/server programming is used. The energy

☆ More information about energy information systems, as well as links to many of the tools discussed in this entry, can be found at www.utilityreporting.com

Keywords: Energy information system; Internet; Meters; Utility data; Web publishing.

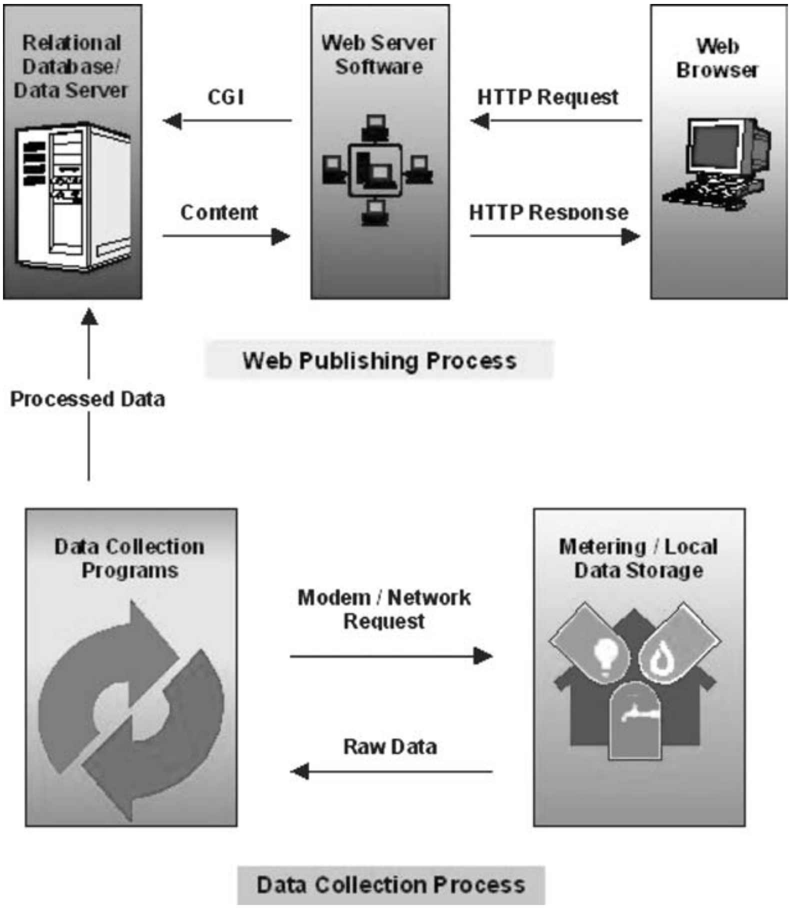


Fig. 1 EIS schematic.

data are stored on a central computer—the server—and wait passively until a user makes a request for information using a Web browser—the client. A Web-publishing program retrieves the information from a relational database and sends it to the Web server, which then sends it to the client Web browser that requested the information.

Many software choices are available for the Web-publishing process. One choice uses a server-side Common Gateway Interface (CGI) program to coordinate the activity between the Web server and the Web-publishing program. CGI is a method used to run conventional programs through a Web browser.The Web-publishing client/server process for an EIS uses the steps below (see Fig. 1):

1. A user requests energy information by using a Web browser (client) to send an hypertext transfer protocol (HTTP) request to the Web server.
2. The Web server activates the CGI interface program, which then starts the Web-publishing program.
3. The Web-publishing program retrieves the information from the relational database, formats the data in hypertext markup language (HTML), and returns it to the CGI interface program.
4. The CGI interface program sends the data as HTML browser content to the Web server, which sends the content to the Web browser that requested the information.

This entire process may take only seconds, depending on the speed of the client computer's connectoin to the Web.

PROGRAMMING CHOICES FOR EIS WEB PUBLISHING

Server-Side Programs

Server-side programs are programs that run on the network server. CGI programming is the most fundamental way to access relational database information over the Internet or an Intranet and to display dynamic content. This is important to any EIS because the amount of data involved will undoubtedly require processing with a relational database of some kind.

FoxWeb is a small CGI program that connects to a Visual FoxPro database application. The application then calls any number of custom-designed queries and procedures to return results to the browser as HTML. More information about FoxWeb is available at www.foxweb.com

Perl (practical extraction and report language) is an application used for CGI programming. Applications written in Perl will work on any operating system. It also has the ability to connect to many types of databases. A good source for information on Perl is www.perl.com

ColdFusion is a server-side application that uses a ColdFusion server. The server executes templates containing a mixture of HTML and ColdFusion instructions and then returns the results to the browser as pure HTML.

Active Server Pages (ASP) are also a popular choice recently. The ASP program on the Windows Web server will automatically interpret Web pages with the extension.asp. The Web pages are a mixture of ASP instructions, Visual Basic code, and HTML.

Java Servlets are Java programs that run on a Web server and build Web pages. Java is the latest of a long line of higher-level programming languages such as Fortran, Pascal, and C++. It is also portable across operating systems. Java Servlets written for one Web server on one operating system will run on virtually any Web server and on any operating system.

Java Server Pages (JSP) are similar to ASP except that the pages consist of Java code instead of Visual Basic code. This makes the code portable across operating systems. The Web pages typically have the extension.jsp. This tells the Web server to interpret the embedded Java code like any other Java Servlet. Information about Java Servlets and JSP is available on this Johns Hopkins University Web site: www.apl.jhu.edu/~hall/java

Hypertext Preprocessor (PHP). In an HTML document, PHP script (similar syntax to that of Perl or C) is enclosed within special PHP tags. Because PHP is embedded within tags, the author can jump between HTML and PHP (similar to ASP and ColdFusion) instead of having to rely on heavy amounts of code to output HTML.

Any organization wishing to develop an EIS should carefully consider which server-side applications to use. The decision should be a practical one rather than a popular one. All of the criteria below should be part of the evaluation process:

- What operating system is predominantly available to the facility?
- What programming languages are support personnel willing to work with?
- What applications are compatible with the existing database?
- How much of the budget is available to spend?

Client-Side Programs

Client-side applications can create a deeper level of interactivity within Web pages. Scripting languages such as JavaScript and VBScript are less complex versions of other languages, such as Java and Visual Basic, respectively. They reside within the HTML of a Web page and provide a great deal of functionality that HTML itself cannot. Scripts such as these validate input, control the cursor programmatically, and do much more.

Dynamic HTML (DHTML) is the result of scripting languages taking advantage of the extensions common to the latest browsers to make the pages change after they are loaded. A good example of this is a link that changes color when the user places the mouse pointer on it. Much more dramatic effects are possible using DHTML. However, the two most popular browsers—Internet Explorer and Netscape—interpret DHTML differently. Good information about DHTML is available at www.dynamicdrive.com

Cascading Style Sheets (CSS) are special HTML features that allow much more flexibility in formatting elements of a Web page. The ability of CSS to describe the style of an element only once rather than every time you display the element provides a separation of content and presentation. This makes Web pages less complex, much smaller, and therefore faster to load. Beware that CSS is fully supported only in the latest versions of browsers (4.0 and later).

Java applets are small Java programs that are stored on a Web server and called from the HTML in a Web page. Statements in the HTML pass to the applet parameters that affec its functionality. Unlike what it does with Java Servlets, the browser downloads the applet and runs it using the browser's own operating system rather than the operating system of the Web server. Free Java applets are widely available on the Internet. KavaChart applets are very useful for charting trends in data. KavaChart applets are available at www.ve.com

Extensible markup language (XML) is a metalanguage that has many uses. A metalanguage is a language used to explain another language. Extensible markup language organizes data in a predefined format for the main purpose of sharing between or within computer systems. Furthermore, its uses include data organization and transfer, data presentation, data caching, and some that probably have not been invented yet. More information about XML is available at http://xml.com

Developers can use any or all of these to enhance the content of an EIS. Three important points to remember about using client-side applications to enhance browser content are

- Many client-side applications require later versions of browsers to work correctly. Be sure that all your users are using the required browser versions.

- Many client-side applications are available free. Search the Internet before spending resources to develop custom client-side applications.
- Client-side applications will make your Web pages more complex, which adds to development and maintenance costs. Be sure to weigh the benefits of these enhancements against their costs.

Choosing EIS Web-Publishing Tools

We might define tools, in this case, as utility applications that require more configuration effort than programming effort. There are some exceptions, of course. In any case, tools fall into three categories:

- Open-source or free
- Purchased
- Developed

Tools perform such functions as batch emailing, charting, scheduling application run times, and enabling database-to-Web connectivity. The relational database and Web server are also tools.

Batch email applications can be of any category. There are good free ones and purchased ones. Some email servers have batch-processing capability, and some do not. Purchased batch email applications are relatively inexpensive. They have some variations in features, such as whether or not they will send HTML or attachments. They are also easy to develop.

Charting tools are really too complex to warrant developing. There are some very good free and open-source charting tools that have all the features of purchased ones.

The EIS needs scheduling programs to launch data collection applications at predefined times. Some operating systems have scheduling programs built in, but they may be difficult to configure. A purchased version is probably the best choice, because the cost will be quite low.

Some applications do most of the database-to-Web connectivity. These applications either require purchase or (in the case of the free applications) a great deal of programming. The purchased applications are a good choice because much of the error reporting is part of the application. The purchased versions can be very expensive or relatively inexpensive. Each database has its own connectivity options, so much of this decision rests on which relational database is used.

The database used is likely a purchased one or may actually be open-source. Open-source databases like MySQL are competing with the best of the others. Commercial relational database systems range in cost from very inexpensive (Microsoft Access) to very expensive (Oracle).

Open-source Web servers like Apache are available for some operating systems and are very widely used and reliable. Others are free with operating systems such as Microsoft Internet Information Server (IIS). Some are also commercially available for a few hundred dollars. The Web-server choice depends mostly on the operating system of the server itself.

The EIS tools will likely be a mix of open-source, purchased, and developed applications. It is important to consider budget constraints, operating systems, and support when deciding which tools and what types of tools to use. Always plan for compatibility among all the tools the EIS will use.

EIS Choice: Purchase or Do It Yourself

Up to this point in this entry, we have gone "under the hood" and provided details on the processes and software tools needed to create a custom EIS. However, users may not be willing to invest the time and effort required for this do-it-yourself approach.

Numerous companies provide an EIS for an ongoing monthly service fee. The services available range from monthly utility billing data processing and analysis to interval data recorded from submeters for utility cost allocation and load profiling. The advantage of this approach is that the user does not get involved with the details and operation of the EIS, but instead is able to work with the EIS service provider to develop the utility data reports most helpful to the user's operation. An ongoing monthly service fee is a function of the amount of data processed; the more meters or bills processed, the higher the monthly fee. There are additional costs for customizing any reporting from the standard reports already created by the EIS service provider. Ultimately, the monthly cost for the EIS service would have to be justified by the benefit from the EIS reports provided.

The do-it-yourself approach works well for users that either have sufficient IT software personnel in house or that can hire an IT consultant to develop and maintain their own custom EIS. However, having adequate IT resources is only half the battle. The user also needs to create an EIS requirements document that defines what data need to be collected and at what frequency, as well as a list of report outputs. As explained earlier in this entry, the software choices for developing the EIS are numerous, which can result in debates among IT personnel as to which system is best. The best solution will be the one that will result in the desired EIS for the least cost in the least amount of time. The energy manager can help facilitate this decision (and prevent IT gridlock) by estimating the cost, advantages, and disadvantages of each approach.

The decision between the EIS service approach and the do-it-yourself approach will be based on several factors, but it will ultimately be based on the ongoing cost to operate the EIS.

UTILITY REPORT CARDS: EIS EXAMPLE

The Utility Report Cards (URC) program[2] is a Web-based EIS that reports and graphs monthly utility data for schools. The URC was developed and prototyped by the Florida Solar Energy Center (FSEC)[3] using Orange County Public Schools (OCPS)[4] utility data. Each month, the EIS automatically generates a Web-based report and emails it to school staff to examine the school's electricity usage (energy efficiency) and to identify schools with high-energy consumption for further investigation. The easy-to-use Web-style report includes hyperlinks allowing users to (1) drill down into further meter details, (2) display graphs for a 12-month comparison with prior-year data, (3) filter the data to show selected schools, and (4) re-sort the data to rank schools based on the data selected. The URC is also for teachers and students to use as an instructional tool to learn about school energy use as a complement to the energy education materials available through the U.S. Department of Energy's EnergySmart Schools program (ESS).[5] To run the URC, go to www.utilityreportcards.com and click URC Live.

URC Data Collection

On a monthly basis, the OCPS utilities (Orlando Utilities Commission[6] and Progress Energy[7]) electronically transmit to FSEC the OCPS utility data, which then adds it to the URC relational database. Because there was no consistency between the utility's data format, FSEC created a custom program called URC_DPP (URC Data Processing Program) that processes each utility's data file separately and loads the data into a common Oracle database.

Florida Solar Energy Center sends an email to a designated email address at OCPS, with a copy of the current month's URC embedded in the email message. Then OCPS forwards this email to its internal email distribution list for the OCPS staff. This procedure makes it easy for users, because all they need to do is click the hyperlinks in the URC email to produce graphs and detailed reports.

URC Web-Publishing Program

The URC program is informative, intuitive, and flexible for all users. It takes advantage of extensive use of hyperlinks that create graphs and detailed reports from an overall summary report listing all schools. Users are able to view graphs and see the electric consumption patterns simply by clicking hyperlinks in the URC Web page.

When the user selects the URC Live hyperlink, a summary report (see the example in Fig. 2) shows totals for each school type in the entire school district—primary (elementary) schools, middle schools, high schools, etc.—as the rows in the report. The data presented in the columns include the electric consumption (kWh), the cost (dollars), and the efficiency (Btu/ft^2). To provide meaningful comparisons to the same time in the prior year, the URC program divides the data by the number of days in the billing period to produce per-day values. The URC also allows the user to change the values shown to per-month

UTILITY REPORT CARDS

Electric Per Day — ORANGE COUNTY PUBLIC SCHOOLS — << February 2004 >>

	Consumption (kWh/day)			Cost ($/day)			Efficiency (Btu/sq ft/day)		
School Type	Current Period	Previous Period	Percent Change	Current Period	Previous Period	Percent Change	Current Period	Previous Period	Percent Change
Regular Primary School	178,073	181,162	-2 %	$ 13,923	$ 12,012	16 %	146	149	-2 %
Regular Middle School	121,735	124,550	-2 %	$ 9,119	$ 7,984	14 %	139	142	-2 %
Regular High School	168,196	167,136	1 %	$ 12,135	$ 10,121	20 %	161	160	1 %
Special Other	6,351	6,092	4 %	$ 479	$ 404	19 %	143	137	4 %
Vocational High School	5,910	6,572	-10 %	$ 469	$ 448	5 %	106	118	-10 %
Grand Total ORANGE COUNTY PUBLIC SCHOOLS	480,266	485,511	-1 %	$ 36,125	$ 30,969	17 %	148	150	-1 %

On — Denotes Increase From Previous Year

On — Denotes Decrease From Previous Year by ◀ 0 % ▶ or more

Fig. 2 Overall school-district summary page.

figures (the per-day values multiplied by 30 days). Fig. 2 shows the overall school-district summary report for February 2004.

The URC program interface makes the program easy to use, considering the enormous amount of data available. The top-down approach lets users view different levels of data from the overall districtwide summary to individual meters in a single school. For example, the user can click a school type to display the details for each school. Fig. 3 shows the result of clicking the hyperlink Regular High School: A Report on All High Schools in the Database for February 2004. The report is sorted based on the percent change in kilowatt-hour usage from the prior-year levels. The schools that changed the most are at the top of the list. Re-sorting the schools is accomplished simply by clicking the column title. To produce the report shown in Fig. 3, the original report sorted by the efficiency percentage change was re-sorted by clicking the consumption percentage change column.

Graphing is accomplished by clicking any current period value in the report. To display a 12-month graph for Apopka Senior High School, clicking the number 22,399 would produce the graph shown in Fig. 4. Note that the consumption levels are significantly higher than prior-year levels for this school. Focusing on the reasons for this increase should be the next step for the OCPS facility personnel. When it is determined that the increase is not due to other factors, such as increased enrollment or extreme weather, personnel can consider making adjustments to the energy management system controls to turn the trend around.

URC Web-Program Details

Two Web programs make up the URC application interface. One is for reporting, and one is for graphing. The programs are written in PHP. "Hypertext preprocessor is a widely-used Open Source general-purpose scripting language that is especially suited for Web development and can be embedded into HTML," according to PHPBuilder.com. This means that the PHP application itself is not for sale. It is developed and supported solely by volunteers. What makes it an attractive choice is that the developers' code is part of the HTML page itself. Hypertext preprocessor is ideal for connecting to databases and running SQL queries return dynamic content to a Web page.[8]

Using PHP, the reports accommodate both novice and expert users. Hyperlinks in the reports pass values back into the same two programs in a recursive manner, using

UTILITY REPORT CARDS

Electric Per Day — ORANGE COUNTY PUBLIC SCHOOLS — << February 2004 >>

Regular High School	Consumption (kWh/day)			Cost ($/day)			Efficiency (Btu/sq ft/day)		
SCHOOL	Current Period	Previous Period	Percent Change	Current Period	Previous Period	Percent Change	Current Period	Previous Period	Percent Change
APOPKA SENIOR HIGH SCHOOL	22,399	20,360	10 %	$ 1,777	$ 1,375	29 %	211	192	10 %
UNIVERSITY HIGH SCHOOL	20,713	19,056	9 %	$ 1,482	$ 1,131	31 %	163	150	9 %
WINTER PARK HIGH SCHOOL	22,631	21,157	7 %	$ 1,604	$ 1,247	29 %	168	157	7 %
EVANS HIGH SCHOOL	17,108	16,526	4 %	$ 1,229	$ 1,022	20 %	156	151	4 %
OLYMPIA HIGH SCHOOL (FORMERLY DR. PHILL	13,441	13,119	2 %	$ 980	$ 805	22 %	118	116	2 %
TIMBER CREEK HIGH SCHOOL	15,133	15,379	-2 %	$ 1,081	$ 924	17 %	134	136	-2 %
WEST ORANGE HIGH SCHOOL	21,375	21,898	-2 %	$ 1,473	$ 1,261	17 %	190	194	-2 %
ROBERT HUNGERFORD PREPARATORY HIGH SCHOOL (FORMERL	5,525	5,727	-4 %	$ 424	$ 393	8 %	192	199	-4 %
OAK RIDGE HIGH SCHOOL	12,940	14,669	-12 %	$ 908	$ 855	6 %	148	168	-12 %
CYPRESS CREEK SENIOR HIGH SCHOOL	16,932	19,245	-12 %	$ 1,176	$ 1,108	6 %	153	174	-12 %
Regular High School	**168,196**	**167,136**	**1 %**	**$ 12,135**	**$ 10,121**	**20 %**	**161**	**160**	**1 %**

Fig. 3 High-school summary report sorted by percentage change in consumption from prior year.

Energy
Eff–Serv

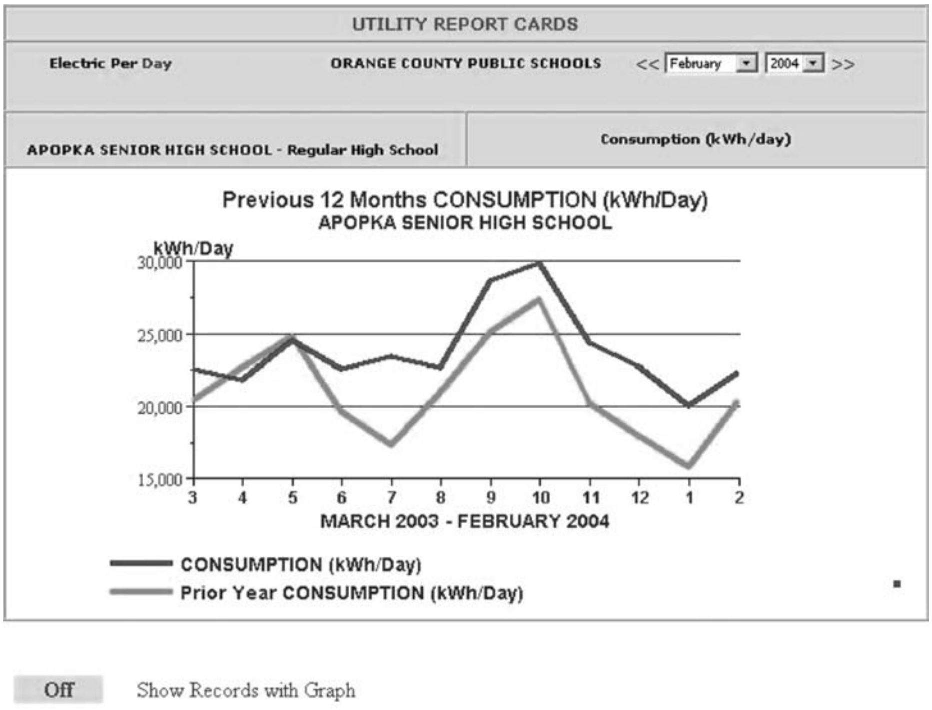


Fig. 4 Apopka Senior High School 12-month kWh graph.

the CGI query string. These values are processed by a set of predetermined rules built into the program by the developer to change the appearance of the reports incrementally to suit the user. The hyperlink construction includes messages using the onMouseOver event to explain the action of the hyperlink. Total and subtotal lines provide summary information. Data that show an increase or decrease from the previous year are flagged with a different background cell color. The user can define percentage criteria for marking decreases from the prior year because this is helpful in tracking progress toward a particular goal, such as a 5% decrease from the previous year. Graphs are created using KavaChart, a collection of Java applets available from Visual Engineering at www.ve.com.[9]

CONCLUSION

Today's energy manager needs to be knowledgeable about the basic principles and concepts of IT because it is a fast-growing area of new systems and services. Web-based EISs provide users feedback on how and where energy is consumed. Building automation systems control the devices that use energy. Together, these two systems provide the energy manager the tools needed for a successful energy management program.

REFERENCES

1. Capehart, B.; Allen, P.; Pawlik, K.; Green, D. How a web-based energy information system works. *Information Technology for Energy Managers*; The Fairmont Press Inc.: Lilburn, Georgia; 2004.
2. Allen, P.; Green, D.; Kalaghchy, S.; Kivler, B.; Sheinkopf, B. The utility report cards—an energy information system for orange county public schools. *Information Technology for Energy Managers*; The Fairmont Press Inc.: Lilburn, Georgia; 2005; Vol. II.
3. Florida Solar Energy Center; A Research Institute of the University of Central Florida; http://www.fsec.ucf.edu/ (accessed April 2004).
4. Orange County Public Schools, Orlando, Florida; http://www.ocps.k12.fl.us/ (accessed 28th March 2007).
5. U.S. Department of Energy's EnergySmart Schools; Rebuild America, Helping Schools Make Smart Choices About Energy; http://www.rebuild.org/sectors/ess/index.asp (accessed April 2004).
6. Orlando Utilities Commission; OUC The Reliable One; http://www.ouc.com/ (accessed April 2004).
7. Progress Energy; People, Performance, Excellence; http://www.progress-energy.com (accessed April 2004).
8. The PHP Group; PHP Manual, Preface; http://www.phpbuilder.com/manual/preface.php (accessed April 2004); last updated, October 26, 2002.
9. Visual Engineering; KavaChart, The Complete Solution for Java-Based Charting; http://www.ve.com/ (accessed April 2004).

BIBLIOGRAPHY

1. Apache HTTP Server Project, Apache Software Foundation, www.apache.org
2. ColdFusion MX, Macromedia, Inc., www.macromedia.com/software/coldfusion
3. Acquisuite Energy Information Server, Obvius, Inc., Portland, Oregon, www.obvius.com
4. FoxWeb, Eon Technologies, Alameda, California, www.foxweb.com
5. Java, Sun Microsystems, Inc., http://java.sun.com
6. KavaChart (Java applets for graphing data), www.ve.com
7. MySQL, MySQL AB, www.mysql.com
8. Perl, Perl Mongers—The Perl Advocacy People, www.perl.org
9. Visual FoxPro, Microsoft Corporation, http://msdn.microsoft.com/vfoxpro
10. PHP, PHP Builder, http://phpbuilder.com
11. Utility Report Cards, Florida Solar Energy Center, www.utilityreportcards.com

Energy Management: Organizational Aptitude Self-Test

Christopher Russell
Energy Pathfinder Management Consulting, LLC, Baltimore, Maryland, U.S.A.

Abstract
Human, technical, and financial criteria all contribute to a manufacturer's ability to build wealth through energy management. Collectively, these attributes constitute a "culture" and receptiveness not only to energy management, but to operational efficiency in general. Manufacturers will enjoy a wider range of energy management options by nurturing several key organizational attributes, including staff awareness, competence, leadership, commitment, and removal of institutional barriers. This entry offers a typology and self test of corporate "aptitudes" for energy management. This discussion is based on the author's observation and research (The author served as director, industry sector, for the Alliance to Save Energy in Washington, DC from 1999 to 2006. His work included (1) over 40 workshops serving industrial energy users, each attracting 25 to over 120 participants; (2) presentations at 30 industry conferences; and (3) over 30 articles for trade press. All of these activities have generated communications with hundreds of individuals, all of whom add another dimension to the cumulative story of industrial energy management). Until these theories can be properly tested, readers are asked to merely consider this persuasive argument.

LIST OF ATTRIBUTES THAT FACILITATE CORPORATE-WIDE ENERGY MANAGEMENT

Fundamental business viability. The manufacturer's front office stability is important. Companies that are the subject of a merger or acquisition, labor disputes, bankruptcy, or severe retrenchment may have fundamental distractions that will interfere with the attention that energy management deserves. A preponderance of such conditions indicates management turmoil that makes energy management impractical.

Replication capacity. Manufacturers with multiple facilities should spread knowledge of energy efficient techniques and compare their ongoing results. The ability to cooperate across multiple sites and across departmental boundaries is required to maximize industrial energy management potential.

Energy leadership (or "champion"). Successful energy improvements are usually led by an "energy champion," a manager that (i) understands both engineering and financial principles (ii) communicates effectively both on the plant floor and in the boardroom, and (iii) is empowered to give direction and monitor results.

Energy market capability. This dimension is straightforward: does the corporation wish to purchase energy through open-market activity or procure as usual from the local utility? If open markets are the choice, the corporation should be prepared to maintain sophisticated search and verification procedures to support its contracting activities. Purchasing decisions should reflect the collaboration of procurement, production, and maintenance personnel.

Leadership intensity. Quality of operations should be demanded, facilitated, and recognized by top officers of the corporation. Adoption of professional and industry standards, such as ISO 9000, are helpful in attaining this attribute. Energy-smart operations will hold employees accountable for adherence to energy management goals and other quality standards.

Pride intensity. Energy efficiency is very much dependent on the behavior of line workers. Employees' awareness of their impact on energy costs must be achieved. A positive, can-do attitude on the part of staff is helpful in attaining potential energy savings. Rewards and recognition can be harnessed to good effect.

Fiscal protocol. Financial considerations involve far more than invoice quotes. Are purchasing decisions made on first-cost or lifecycle costs? Which department pays for improvements and which claims the savings? Do savings count only fuel bill impacts or do they include the value of material waste minimization and greater capacity utilization? What criteria determine adequate payback?

Engineering protocol. Successful energy management depends on an ability to understand energy consumption. This requires benchmarking, documenting, comparing, remediating, and duplicating success stories. Internal skills, procedures, and information services are engaged. The likelihood of building value through energy efficiency varies directly with the depth of these technical capabilities.

In the absence of an energy management process, energy expense control is reduced to one-dimensional efforts. Many manufacturers (either wittingly or not) settle for something less than full energy efficiency potential due to a lack of time, interest, or understanding. The approach

Keywords: Energy; Energy efficiency; Energy management; Industry; Manufacturing; Risk management.

taken by individual manufacturers is very much a function of their organizational attributes and business culture.

PREVAILING ENERGY MANAGEMENT STRATEGIES

The aim of this section is to present the range of typical energy management strategies practiced by industry. Every manufacturer employs some energy management strategy, even if the choice is to do nothing about energy consumption. Consequently, every manufacturing organization adopts one or more of these strategies:

- *Do Nothing*. Ignore energy improvement. Just pay the bill on time. Operations are business-as-usual or "that's the way we've always done it." The result is essentially "crisis management" in that energy solutions are induced by fire-drill emergencies and undertaken without proper consideration of the true costs and long-term impacts.

  Who does this? Companies that do not understand that energy management is a strategy for boosting productivity and creating value. Or, companies that are subject to merger, buy-out, bankruptcy, union disputes, relocation, or potential closure. Or, companies that are extremely profitable and don't consider energy costs to be a problem.
  PROs: you don't have to change behavior or put any time or money into energy management.
  CONs: you don't save anything. Income is increasingly lost to uncontrolled waste. Because you don't inventory your energy usage, you are exposed to volatility in energy markets. You are less prepared to adapt to evolving emissions compliance agendas and you are less capable of spotting opportunities presented by new technologies. Because you don't monitor anomalies in energy flow data, you are more susceptible to lapses in mechanical integrity and plant reliability.

- *Price Shopping*. Switch fuels and shop for lowest fuel prices. No effort to upgrade or improve equipment. No effort to add energy-smart behavior to standard operating procedures.

  Who does this? Companies that "don't have time" or "don't have the money" to pursue improvement projects. Or, these companies truly believe that fuel price is the only variable in controlling energy expense.
  PROs: you don't have to bother plant staff with behavioral changes or create any more work in the form of data collection and analysis.
  CONs: lack of energy consumption knowledge exposes the subject company to a variety of energy market risks. You don't know where your waste occurs nor do you identify opportunities to boost savings and productivity. You are also exposed to energy market volatility and emissions and safety compliance risks.

- *Occasional O&M Projects*. Make a one-time effort to tune-up current equipment, fix leaks, clean heat exchangers, etc. Unable/unwilling to make capital investments. Revert to business-as-usual O&M behavior after one-time projects are completed.

  Who does this? Companies that are insufficiently organized to initiate procedural changes or make nonprocess asset investments. They cannot assign roles and accountabilities for pursuing ongoing energy management.
  PROs: you spend very little money when just pursuing quick, easy projects.
  CONs: savings are modest and temporary because you don't develop procedures for sustaining and replicating your improvements. Familiar energy problems begin to reappear. Energy bills begin to creep back up.

- *Capital Projects*. Acquire big-ticket assets that bring strategic cost savings, but beyond that, day-to-day O&M procedures and behavior are business-as-usual.

  Who does this? This strategy is adopted by companies that believe that advanced hardware is the only way to obtain real, measurable savings. Similarly, they believe that operational and behavioral savings are "weak" and not measurable. However, they have the fiscal flexibility to acquire strategic assets that boost productivity and energy savings.
  PROs: obtain fair to good savings without having to change behavior or organize a lot of people.
  CONs: forfeit savings attributable to sustained procedural and behavioral efforts. Also, savings from the new assets may be at risk if adequate maintenance is not applied.

- *Sustained Energy Management*. Merge energy management with day-to-day O&M discipline. Diagnose improvement opportunities and pursue these in stages. Procedures and performance metrics drive improvement cycles over time.

  Who does this? Companies with corporate commitment to quality control and continual improvement, well-established engineering and internal communications protocol, and staff engagement through roles and accountabilities.
  PROs: maximize savings and capacity utilization. Increased knowledge of in-plant energy use is a hedge against operating risks. Greater use of operating metrics will also improve productivity and scrap rates while reducing idle resource costs.

CONs: you need a lot of in-house talent, cooperation, and a capable energy "champion" to do this.

It is beyond the scope of this entry to comment on which strategies are predominantly encountered in industry. Anecdotal evidence suggests that all industrial energy management strategies can be categorized per one of these five selections. It is also possible for firms to practice multiple strategies simultaneously—for example, price shopping for low-priced fuel commodities in concert with a capital-projects focus.

It should be noted that most of the ten of the experiences documented in the Alliance's corporate energy management case study series can be categorized as "sustained energy management." As such, these companies integrate energy management with day-to-day operating procedures and accountabilities.

ENERGY MANAGEMENT PATHFINDING: MATCHING STRATEGIES WITH CORPORATE ATTRIBUTES

This section will build on the theory of corporate receptiveness to energy management, as presented above. The energy management strategies available to a manufacturer are a function of its organizational attributes, as summarized in Table 1. Note that this is currently presented as theory.

Examples for interpreting this Table 1 manufacturer should have attained the attributes of "fundamental viability," "leadership intensity," "fiscal protocol," and "engineering protocol" in order to effectively pursue capital projects as a single-site energy reduction strategy.

Alternatively, a manufacturer that has attained "fundamental viability," "replication capacity," "leadership intensity," "pride intensity," "engineering protocol," and has an "energy champion," should be capable of pursuing both the occasional O&M projects and sustained energy management strategies across multiple sites. In this instance, the company may wish to start with the lesser strategy (O&M projects) and evolve into the practice of sustained energy management.

This typology presumes that energy management for multisite organizations is more demanding than it is for single-site companies. Accordingly, adoption of a certain strategy by a multisite organization requires all the organizational attributes that a single-site organization would be expected to muster, plus the capacity to replicate.

Managers that are contemplating improved energy management are encouraged to consider the case study results and theory presented in this paper. To act on this information, the steps are to:

Table 1 Theory: matching corporate attributes to energy management strategies

	Organizational attributes							
	Fundamental viability	**Replication capacity**	**Energy champion**	**Energy market capability**	**Leadership intensity**	**Pride intensity**	**Fiscal protocol**	**Engineering protocol**
Strategies for single-site energy reduction								
Do nothing								
Price shop				Required				
Capital projects	Required				Required		Required	Required
Occasional O&M projects	Required				Required	Required		
Sustained energy management	Required		Required		Required	Required		Required
Strategies for replicating energy reduction at multiple sites								
Do nothing								
Price shop		Required		Required				
Capital projects	Required	Required	Required		Required		Required	Required
Occasional O&M projects	Required	Required	Required		Required	Required		
Sustained energy management	Required	Required	Required		Required	Required		Required

Source: From The alliance to save energy.

- Refer to Appendix A, "Determining an Organization's Aptitude for Energy Management." Note which organizational attributes have been substantially attained by the subject company.
- Compare the attained attributes to the information in Table 1. The presence (or absence) of certain attributes determines which energy management strategies are available to the subject company.
- Use these findings to understand what the subject organization can or cannot achieve in terms of energy management.

Keep in mind that this exercise indicates what a manufacturer can expect from energy management given its current organizational attributes and business culture. There may be a desire to evolve to a higher level of energy management than what the current organization allows. What if a manager wants to advance energy management in his or her organization? There are windows of opportunity. An obvious example is when energy market turmoil brings top management's attention to fuel costs. Also, take advantage of annual planning sessions or strategic reorganizations to propose the kind of organizational processes needed to practice sustained energy management. Remember that energy cost control is as much dependent on people as it is on technology.

CONCLUSION

Volatile energy markets are here to stay, and so are competitive and regulatory pressures. Energy price movements will put some manufacturers out of business, while others will decide to move offshore. Surviving manufacturers will not only provide superior products and service, but they will maximize value through operating efficiencies. Energy efficiency is an indispensable component of wealth creation.

Energy procurement strategies such as shopping for low energy prices and supply contracts are only partial solutions to soaring energy expenses. Management of consumption is an underappreciated opportunity. While technology is the foundation for managing consumption, it is the human dimension that makes technology work. Organizational procedures, priorities, and accountabilities are crucial to energy management.

A manufacturer's ability to manage energy consumption is ultimately a function of organizational attributes and corporate culture. This entry advances "energy management pathfinding" concepts. Appendix A presents the criteria that define seven distinct organizational attributes needed for energy management. While sustained, day-to-day energy management is recommended for providing the greatest and most durable value, and it is also the most demanding in terms of operational character. Many companies will find that they are suited for strategies that are less challenging but may also provide less value. The same management diagnostic presented in this entry serves as a pathfinder for matching organizational characteristics with appropriate energy management strategies.

APPENDIX A

Determining an Organization's Aptitude for Energy Management

This Appendix serves two purposes:

- To further define the organizational attributes that a manufacturer needs to pursue energy management as a continuous-improvement process, and
- To determine if a subject organization has substantially attained each of the organizational attributes listed ("Fundamental Viability," "Replication Capacity," etc.).

Please see below. For each attribute, a number of conditions are posed in a bulleted list. When considering a subject company, ask: are most or all of these conditions true? If the answer is yes, then the subject company has substantially attained that attribute. The degree of attainment for each attribute varies directly with the number of considerations that can be affirmed for each attribute. There are no scores, per se. If the subject company has attained a majority of the bulleted considerations listed under an attribute, consider that attribute to be substantially attained.

The range of topics covered by these conditions would be best answered by a high-level manager or perhaps a team of managers. After this exercise, note all of the attributes that have been substantially attained. Compare those results to Table 1. That table indicates which energy management strategies are available to the company, given its organizational attributes.

Fundamental Viability

Your plant capacity is generally stable or growing.
Your company is not currently experiencing excessive turnover of managerial and corporate personnel.
Strikes or other labor-related work stoppages are not considered an ongoing concern for management.
Your company is not the current subject of a merger or acquisition attempt.
Your company is not in receivership, Chapter 7, or Chapter 11 status.

Replication Capacity

Your company operates more than one manufacturing facility.

Your manufacturing processes and products are mostly similar across all plants.
Your facilities are designed and operated per one standard; standards do not significantly vary by facility for asset selection, procedures, and management styles.
Staff members from different plants (or divisions) regularly collaborate to share their common issues and solutions.
Maintenance management is set up to serve multiple sites; individual sites adhere to centralized maintenance planning and procedures.
Your corporation currently uses (or is it willing to use) contract vendors for ongoing energy management.

Energy Champion (Note: All of These Conditions Must be Met to Have a True "Energy Champion")

Your lead energy person has thorough knowledge of technology and staff capabilities at the facility level.
Your lead energy person can prepare financial analyses to support engineering proposals and convincingly present these to top managers.
Your lead energy person applies more than 50% of his or her time to energy issues.
Your lead energy person can give direction or at least influence decision making by general managers.
Your lead energy person understands utility tariff structures and administers relations with utility providers.

Leadership Intensity

Your organization actively maintains disciplines of excellence such as Six Sigma, ISO 9000, or Total Quality Management.
Process technologies, procedures, or staff expertise are a selling point in marketing your products.
Current and future environmental impacts from manufacturing operations are a concern to your top management.
A corporate officer consistently reviews cost and quality performance data for all facilities.
To most of your corporate leaders, "energy efficiency" is perceived as an "opportunity" as opposed to a "hassle".
Staff compensation, raises, and rewards are impacted by their stewardship of energy, raw materials, and other inputs.
Production metrics are integral to performance evaluations for facility managers and staff.
Your facilities are subject to public scrutiny or "good citizenship" expectations.

Pride Intensity

All or most plants are consistently high performers with respect to health and safety compliance.
Most plant-floor staff members are well trained for their jobs.
Staff turnover is not considered to be a problem.
Your typical plant worker philosophy can be described as "do what's right" instead of "do what's easy."
You describe your plant equipment as "well maintained" as opposed to "poorly maintained."
To most of your facility staff, "energy efficiency" means "opportunity" as opposed to "hassle."
Key facility personnel maintain professional certifications.
Your organization prescribes and enforces technical training for facility personnel.

Fiscal Protocol Intensity

Asset purchases are judged primarily by life-cycle costs (acquisition plus life-time operating, maintenance, etc.), instead of first costs (cost of acquisition).
Your organization uses (or is willing to use) leases and other off-balance sheet methods to finance major acquisitions.
Your organization's investing strategy seeks large payback as opposed to fast payback.
Most of your facilities take utility tariffs into account when planning their operating times.
Facilities invest in plant improvements (as opposed to simply fixing what's broken).
Energy-related capital project proposals assigned a hurdle rate equal to or lower than other project proposals.
Any energy savings are returned to the facilities that successfully implement capital improvements.
Your facility managers understand utility tariffs and their role in determining energy expenses.

Energy Market Capability

Your company is willing to make an on-going effort to use energy marketing services to obtain lowest-cost energy commodities and risk-hedging securities.

Engineering Protocol Intensity

Your facilities maintain a scheduled maintenance routine for powerhouses, motor drives, pumps, compressed air, and similar utilities.
Your facilities maintain a protocol for responding to anomalies in operating performance data.
Your chief engineers are comfortable with using software to analyze engineering issues.

Plant managers develop (or help to develop) project proposals for capital budgeting purposes.
Your facilities maintain procedures for safety, health, and waste management.
Most or all of your facilities maintain an action plan for improving process efficiencies.
Your organization maintains a database or archive that documents engineering problems and solutions.
Your facilities track the volume of factor inputs required per unit of production.
Your facilities monitor scrap or error rates.
Your annual budgets include factor inputs and production targets as well as dollar figures.
Production, inputs, and cost-performance data are created and utilized at the facility level.
Your engineering problems and emergencies are generally unpredictable and unique as opposed to predictable and recurring.
Company-wide production stats are made available to all facility staff by publication, discussion, or graphic display.

BIBLIOGRAPHY

1. Thumann, A. *Plant Engineer & Managers Guide to Energy Conservation*, 6th Ed.; The Fairmont Press: Lilburn, GA.
2. Curl, R.S. *Successful Industrial Energy Reduction Programs*, The Fairmont Press: Lilburn, GA.
3. John, M.S. *Slashing Utility Costs Handbook*, The Fairmont Press: Lilburn, GA.

Energy Master Planning

Fredric S. Goldner
Energy Management & Research Associates, East Meadow, New York, U.S.A.

Abstract
Energy master planning (EMP) is the process of transitioning an organization's culture from the traditional "fixed cost, line item" view of energy to one in which energy is recognized as both an opportunity and a risk that that can be managed. An EMP can guide an organization in longer-range planning of energy cost reduction and control as part of facility maintenance, management, and design. Though an EMP encompasses traditional efforts to cut energy costs, it also includes many steps not usually taken under conventional, technically oriented energy management, including energy procurement, energy-related equipment purchasing, measurement and verification (M&V), staffing and training, communications, and setting energy consumption targets and tracking/feedback loop systems. A critical difference between energy master planning and traditional energy management is an orientation toward the future. An EMP deals with a longer timeframe than just simple payback periods and goes beyond merely reducing energy use. EMP requires sustained commitment from top levels of an organization down through the rank and file, a dedicated energy manager, and creation of an energy team with membership from across the organization.

INTRODUCTION

Developing an energy management program or more broadly energy master planning (EMP) is the process of transitioning an organization's culture from the traditional "fixed cost, line item" view of energy to one in which energy is recognized as the opportunity and risk that it has become.

An EMP can guide an organization in longer-range planning of energy cost reduction and control as part of their facility maintenance, management, and design. An EMP can even lead the energy budget to be recognized as a potential profit center and source of opportunity rather than just another business expense. An EMP moves beyond the confines of traditional engineering to include energy procurement, energy-related equipment purchasing, measurement and verification (M&V), staffing and training, communications, and setting energy consumption targets and tracking/feedback loop systems. The long-term perspective goes beyond simply cutting last year's energy use. It makes energy awareness part of the everyday operation and "mindset" of the organization.

If you're thinking this doesn't apply to your firm or clients because you're too small, think again. This approach works for an organization as small as a single site to an owner with half a dozen small buildings to Fortune 100 companies. The effort and level of detail vary, respectively, but the approach is basically the same. The good news is that there are resources available to assist professionals and the organizations they serve to understand the EMP process and get started on this path.

To be successful an organization must treat energy in the same business-like manner that they do all other major expenses, such as labor and materials. "If you can measure it, you can manage it"[1] is the catch phrase of Paul Allen, the energy manager at Disney World, who has been instrumental in implementing one of the most successful EMPs in the country. "Energy is a competitive opportunity… Winners manage it effectively!" is the driving force of another highly successful program at Owens Corning. Many other organizations throughout this country and across the globe have recognized that to achieve significant and sustained energy cost control, organization needs to make energy management an integrated part of their business/operations. There are various ways to pursue the process, but a key requirement is an interdisciplinary mix of engineering/technical, behavioral, and organizational or management components. The EMP must be integrated into the basic business operations.

Keywords: Strategic; Energy management; Bottom line; Upper management; Profit center; Buy-in; Commitment; Long-term; Integrated; Energy accounting; Business planning; Road map.

UNEXPECTED BENEFITS

One of the most potent driving factors in many organizations' efforts to address energy issues is increased profitability that can be realized through reduced/optimized energy expenditures. Beyond the "bottom line" impacts, an EMP can also provide an organization with a more secure energy supply, reduced downtime of systems, improved equipment availability, reduction in maintenance costs/premature system replacement expenditures, and overall productivity gains. Additional benefits that have been documented include quality-of-life improvements, enhanced product quality, better operational safety, reduced raw material waste in industrial plants, and increased rentability in commercial facilities. An often overlooked outcome of an EMP is the reduction of emissions, among other environmental impacts that help organizations become perceived as better corporate citizens.

As more companies move toward an integrated corporate strategy that links environmental, economic, and social considerations, the results of an EMP can be used to considerable public relations advantage. Ratings in one of the sustainability indices and publication of an annual sustainability report (using, for example, the Global Reporting Initiative guidelines) can give an organization a higher standing in the business community and can result in a higher level of trust by stockholders.

A PROCESS FOR OPTIMIZATION

Though energy prices are volatile, and energy security is often far from reliable, facilities now face leaner operating funds and increased directives to do more with less. Optimizing a facility's operations budget frequently means cutting energy costs. But how do you do so without cutting occupant comfort or productivity? How do you know where to start and what steps to take? And how do you persuade upper management that energy costs can be controlled?

While the general goal of an EMP is the same as that of conventional energy management, the two disciplines are far from identical. Traditional energy management, which is technically oriented, is essentially centered around the boiler or mechanical room. Energy master planning, on the other hand, is a business management procedure for commercial, institutional, and industrial operations. With this approach, it is not enough simply to manage installations. The process involves:

- Developing strategies
- Creating processes to fulfill those strategies
- Identifying barriers and finding procedures to overcome them
- Creating accountability
- Providing feedback loops to monitor and report progress.

Clarification of the terminology for these disciplines is important, as the terms mean different things to different people. In other English-speaking nations, for example, the defining feature of "energy management" is the emphasis on integration with business practices to analyze, manage, and control energy. In the U.S., however, "energy management" has traditionally referred to developing technical and operational measures involving equipment handled by facility managers—not processes such as energy procurement and business planning usually handled by purchasing agents, production personnel, and corporate economists. A typical U.S. energy manager's responsibility rarely extends much beyond utility bill analysis, an occasional energy audit, or managing installation of system upgrades.

Though an EMP encompasses traditional efforts to cut energy costs, it also includes many steps not usually taken under standard energy management. Rather than being just equipment-oriented, an EMP starts long before a comprehensive energy audit and extends beyond commissioning of new systems. An EMP may be thought of as a road map to savings that starts before and continues after energy-efficiency measures are involved. Why do you need such a map? Because you can't get there if you don't know where "there" is. How many of us are willing to undertake a trip that will have costs and risks (like any business or personal decision) if we don't know where we are going? A map makes clear not just your final destination, but also how to get there—and it leaves no question about the starting point.

An EMP is a process to organize and improve your existing energy-related resources and capabilities. Resources, in this case, include standard operating procedures, institutional memory, and actual records (such as energy bills, plans and blueprints, and energy contracts). Capabilities include facilities staff familiar with mechanical room equipment, consultants for energy costs or usage, energy cost accounting and management systems, and meters and software that monitor them. Once organized and integrated, these resources and capabilities become powerful tools for managing energy and producing savings.

TODAY'S PRACTICES FOUND WANTING

Current thinking about managing energy often falls short of the EMP perspective. Today's energy management frequently reflects a short-term crisis mentality: A facility manager or energy manager concentrates on whatever immediate 'fires' have to be put out at his facility. By contrast, with an EMP mindset, the energy manager might first look to increase the efficiency of systems already in place and then move ahead to lay a solid foundation for improving performance via tight energy specs and training.

Current thinking is also often characterized by piecemeal 'solutions' that lead to short-sighted component replacement. When equipment breaks down, it gets replaced without anyone's asking whether this is the best option, long-term. Typically, business thinks in a quarterly mindset because of the short budget cycles of our economic system. That kind of "right now and right here" viewpoint creates situations in which life-cycle thinking is not possible because potentially higher initial costs are visible, but potential benefits tend to be invisible. As a result, when first cost becomes the main criterion for purchasing, such a focus distorts planning and decision-making. Too often, facilities choose the cheapest solution, based on the current quarter's budget, without realizing it could cost them more later.

A critical difference between energy master planning and conventional energy management is an orientation toward the future. With an EMP, you don't just look to increase the efficiency of systems already in place. Instead, you plan for new or changed loads based on your detailed information about the facility's long-term business strategy and projected growth.

Energy master planning deals with a longer timeframe than just simple payback periods. For example, typical financial constraints for a commercial building upgrade often dictate a 2- to 3-year timeframe. Energy master planning, however, looks deeper than simple payback and goes beyond merely reducing energy use. Therefore, an energy professional with an EMP considers life-cycle costs and views long-range planning of energy cost minimization/optimization as part of overall facility maintenance, management, and design. To sustain the savings over time, an EMP calls for hiring an energy manager/coordinator and setting up an energy team. But who is this energy leader? Not simply the facility manager wearing yet another hat, but rather, a highly trained and, often, certified specialist with sufficient acumen and expertise to understand, handle, and maintain whatever new energy systems and practices are to be put in place. Willingness to identify (from within the organization) or hire an individual with the required qualifications and to define his or her responsibility as managing energy rather than managing the facility is a requisite indication of senior management's true commitment to energy master planning. The policy guidance needs to come from senior management levels to convey "buy-in" throughout the management structure of the organization.

IMPROVING BUSINESS AS USUAL

Energy master planning is one way of creating a new norm of "business as usual," taking its cues from time-tested business management practices. For example, rather than carrying out upgrade projects with a defined start and end point, an energy manager uses processes that continue to turn up new sources of savings. This requires a level of creativity for identifying and capturing new opportunities. If a chiller replacement is needed in Building A, for instance, a better approach might be to expand the capacity of the existing central chilled water plant in Building B and run piping from Building B to Building A.

Without an EMP, the Purchasing department often buys equipment and energy, rather than the Facilities department. And Facilities is so busy handling emergencies that they have little interaction with other departments in the organization, let alone industry groups or other end users. An EMP avoids either departmental isolation and turf wars by including representatives from such departments on the energy team. The team consists of more than an energy manager, facilities personnel, and design and construction specialists. To be effective, it needs to incorporate representatives from every department in the organization impacted by energy. This may mean including purchasing, accounting, engineering, environmental affairs, maintenance, legal, health and safety, corporate relations, human resources and training, public relations and marketing, and members from the rank and file (hourly employees).

Energy master planning is a significant challenge. It often rejects the status quo and may question existing components of an organizational culture that do not support energy master planning. If you have always done something one way, you don't necessarily have to perpetuate what could be a costly mistake. For example, using outdated specs (e.g., calling for T12 lamps instead of T8) allows inefficiency to continue.

ORIGINS OF THE AMERICAN EMP APPROACH

A common set of energy master planning definitions and processes has taken root in English-speaking countries other than the US. In the United Kingdom, the energy master planning concept has been practiced for well over a decade and vigorously promoted by the government's Action Energy program. It has been so successful that Canada, New Zealand, and Australia have each adapted the process to their own conditions.

In the US, there are a few recent models cover some, but not all, aspects of energy master planning. management system for energy (MSE) 2000 is a specialized quality improvement standard (like ISO 14000), developed by Georgia Tech. It's an American National Standards Institute (ANSI)-approved management system for energy that covers all sectors, not only buildings. Georgia Tech offers a certificate program to train energy professionals in this standard. The Association of Energy Engineers delivers the Developing an Energy Management Master Plan and Creating a Sustainable Energy Plan Workshops to both 'real' and 'virtual' end users. Live presentations and online seminars present an energy master planning approach with strong emphasis on integration with business strategy.

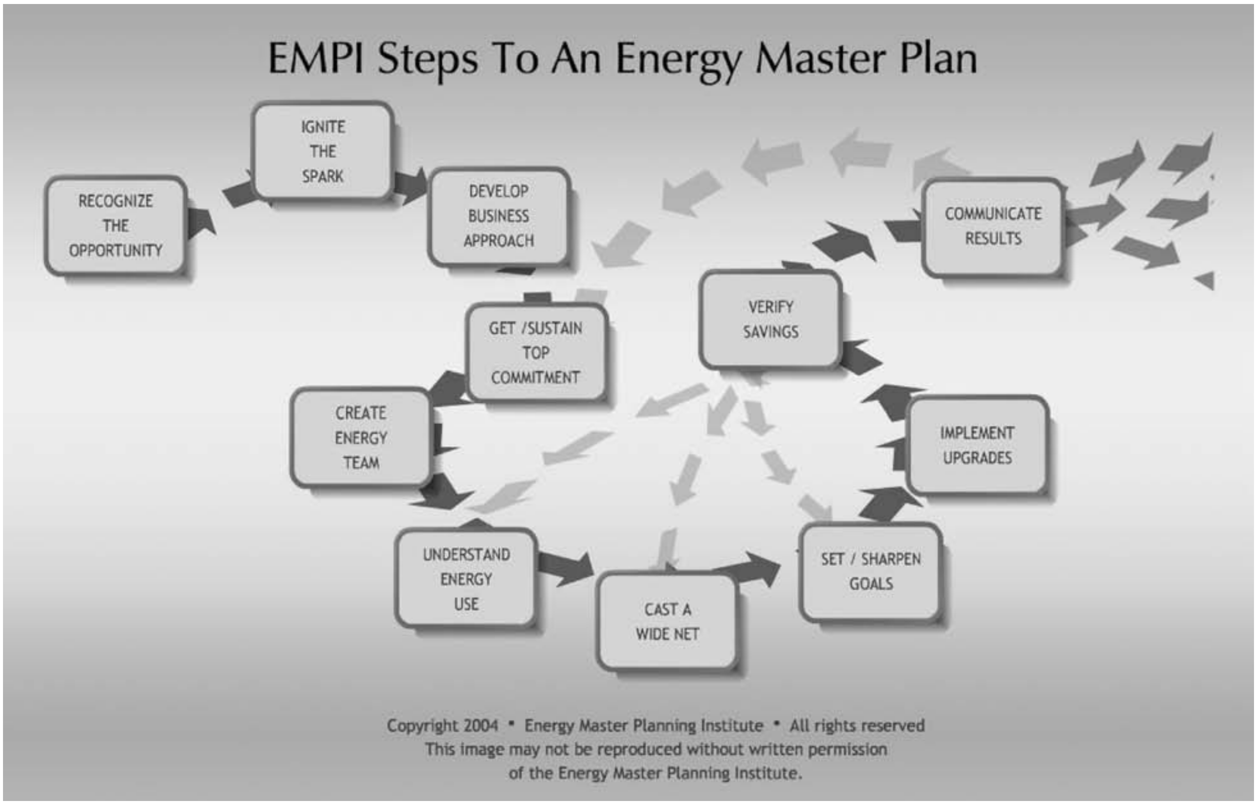


Fig. 1 Energy Master Planning Institute (EMPI) steps to an energy master plan (EMP).

STEPS TO AN ENERGY MASTER PLAN

To fill this gap in the U.S., the Energy Master Planning Institute (EMPI) was established and has developed a set of steps that lay out the process for applying energy master planning to a commercial, institutional, or industrial facility (See Figure 1). This model, which builds on accepted international approaches, offers US organizations a broad and integrated business approach for managing energy that is both strategic and sustainable.

The steps presented in Figure 1 appear sequential, but built into the energy master planning process are series of feedback loops, evident in figure 1. These should not be overlooked, as they guarantee the viability of the process and offer many points for input from internal and external stakeholders, from the Chief Executive Officer (CEO) to boiler room personnel, and from the local community to organizational peers across the country.

Recognize the Opportunity

This is where the process starts—becoming aware of the facility's major energy-related opportunities and challenges. Whether it's the facility manager or the energy manager, whoever takes the initial action must define the opportunities and challenges succinctly so they can be clearly communicated. When the leader creates and implements an EMP, that plan can actually generate a revenue stream and gain recognition for the leader's contribution to the company's bottom line. By integrating energy concerns into the overall corporate business strategy, the energy budget will come to be seen as a potential profit center and source of opportunity—and not just an uncontrollable expense.

Ignite the Spark

Since top executives in an organization are the ultimate decision-makers, particularly concerning funding, it's critical to spark their attention early in the process. It's unlikely that the facility manager (or the energy manager, if there is one), has direct access to upper management, so the right individual to make the pitch to the CEO or Chief Financial Officer CFO must be identified. This could be someone along your management chain, or a consultant or board member or a senior officer in another department who can sell an idea at the top. To find the right person, it's important to understand the decision-making structure of your organization as well as the vision, mission, and long-term business plan.

Develop a Business Approach

To get access to senior executives and persuade them to listen to a new idea, you must speak to them in their own language—dollars/ft^2/year savings, not kWh. Since most executives have not been introduced to the bottom-line value of energy master planning, your task is to change management culture so executives no longer view energy as an uncontrollable expense. Part of the marketing message is that managing energy is no different from tracking, controlling, and accounting for the costs of raw materials, IT, personnel, safety, or the corporate fleet.

Obtain and Sustain Top Commitment

Without question, this step is the most difficult in the energy master planning process. Serious commitment from senior management means providing ongoing financial resources and personnel with appropriate credentials. A one-time memo of support is not effective. To secure top executive commitment, you must show how an EMP can support key business goals, such as growth, customer satisfaction, or a sharper competitive advantage. You might, for instance, explain that lower operating costs and increased energy efficiency can bring a higher level of occupant comfort, which, in turn, can mean a lower worker absentee rate–and, possibly, a greater employee retention rate. To gain top executive commitment, make sure that CEO and CFO see the dollar savings highlighted in the energy team's regular reports.

Top management commitment is the single most important goal in an effective and lasting EMP. Not only is it crucial to have this commitment, it must also be obvious to everyone throughout the organization. Top management should participate in the program start-up and continue to reinforce that commitment periodically with both words and actions. Such organizations as Walt Disney have achieved this top-level buy-in and gained significant and lasting bottom-line results from energy master planning.

Create an Energy Team

A middle manager in your organization may declare that energy master planning is only the responsibility of the Facilities department. However, when an energy team represents the company's broad interests, energy concerns can be successfully integrated into the overall business plan. Every department that's impacted by the organization's energy use should be invited to participate on the team. This includes:

- Facilities
- Construction/engineering
- Purchasing
- Accounting
- Inventory
- Environmental, health, safety
- Legal
- Public affairs
- Property/asset management
- Leasing/real estate
- Risk management
- Security
- Financial service

This should result in the creation of an Energy Committee. Depending on the size of the organization, there may be separate Technical and Steering Energy Committees. In addition to representatives from the departments listed above it is important to include members of the rank & file (hourly employees or line workers). By including such folks in the planning process it allows not only those individuals, but their peers as well, to become aligned with the EMP objectives and particular initiatives during the early stages. Having these employees on the team helps the rest of the rank and file staff (who will be needed to carry out many of the activities) see the energy program as something other than 'just another management flavor of the month'. It also increases success, as feedback from these employees often helps address many of the nuts-and-bolts "bugs" in advance.

The energy team should be headed by an Energy Manager or coordinator. The Energy Manager needs to have a mix of technical, people, and communications skills, and must be enthusiastic. He or she should thoroughly understand the organization's operations and should report to someone as close to the top of the organization as possible, so he or she has the clout need to get things accomplished.

Understand the Organization's Energy Use

Conduct disciplined information analysis. Collect and use metered–not just billed–energy data, and analyze it with software designed to manage energy costs. "Fully understanding current energy use practices–the when, where and how much of energy consumption–in detailed qualitative terms is an essential precondition to energy management master planning."[2] For example, to determine exactly where the energy is going, specify and install sub-metering and a data acquisition system. Such tools will help you monitor and collect, as well as analyze, the meter data for electricity, fossil fuels, steam and condensate, and water. For loads of several hundred kW, it's advantageous to use interval meters that allow a close look at 15-minute interval or hour-by-hour use of energy, so you can see where peaks and valleys in usage occur. This kind of metering is also essential for internal billing. Besides establishing points of excessive usage, it can also help manage loads and pinpoint efficiency opportunities.

Cast a Wide Net

Casting a wide net means looking beyond the central plant for savings opportunities, such as lighting, office equipment, elevators, and localized process loads. It means making construction and equipment specs energy-conscious and creating ways to "enforce" such specs. For example, specifying T8 lamps for efficiency upgrades is not sufficient. The architecture, design, and construction staff should build T8s into their specs for non-energy-related upgrades, such as converting a library to training rooms. Casting a wide net also means revising inventory and purchasing practices to support energy efficiency. As in the example above, to ensure that the new installation continues to function properly, the purchasing specs need to list T8s, not T12s.

Set or Sharpen Goals

Once the organization's energy use is determined, strategic thinking linked to the long-term growth objectives will determine how much energy you should aim to save, where the savings should come from, and when those savings should occur. The energy manager along with the energy team, proposes a phase-one timetable with quantifiable goals to reduce energy use and operating and energy costs. Senior management, however, must mandate the goals and schedule or they carry no weight. Ideally, the CEO, CFO, or the Board issues a position calling for measurable reductions at the end of a 3-year fiscal cycle, based on current use: a percentage reduction in peak electrical demand across the entire facility, a reduction in annual kWh consumption of electricity, and a reduction in overall British Thermal Unit (BTU) for fuel per gross square foot. Lesser annual goals will help keep progress on track. A key is to set achievable goals, and then work to meet or beat them.

Implement Upgrades

Because energy managers generally have experience with efficiency upgrades, it is critical not to fall into old patterns of sporadic efforts and a piecemeal approach, with an eye on the quickest payback. To carry out an EMP, start the upgrade process with a comprehensive energy audit—not just a walk-through—to determine which upgrades will give the best results, not just the best rate-of-return. For example, in a commercial building operation, to avoid a rush upgrade for a new tenant, pre-audit all your buildings so you know what work needs to be done before that tenant signs their lease. Or, if you're considering recommissioning, look at it from the longer-term energy master planning perspective. Recommissioning may not be sufficient, as it means only the existing systems would operate more efficiently. But what if they need to be ripped out as part of the upgrade?

Don't fall into the common trap of focusing just on the high-profile, capital-intensive, projects. While a new chiller or micro-turbine cogeneration system may be a good photo opportunity, focusing on the less visible details can often provide the lion's share of savings through lower-cost measures that improve operations & maintenance (O&M) or maintenance and management practices.

Verify Savings

Once the upgrade process is underway, you must first validate the savings with a recognized technique such as the International Performance Measurement & Verification Protocol (IPMVP or MVP). However, instead of following the M&V protocol after the fact, build measurement and verification into the design—that is, install it as part of the upgrade—so you can identify the savings as soon as the device is turned on. Then, set up an energy accounting system that tracks usage and savings, thereby providing objective accountability. All too often energy management activities are not perceived as being of value because the energy team did not plan ahead and position themselves and theirs efforts for recognizable success. A North American floor coverings manufacturer reported how "in the past we would complete a project to find that we did not have the baseline or operational data to judge whether the project was successful."[3] How willing will management be to fund the next energy related project or initiative if the energy team cannot prove that prior efforts actually saved what was projected?

With an energy accounting system, everyone on the energy team will be working from the same data, and you can prepare regular progress reports, plus quarterly reports to the CFO and an annual report to the Board. This is critical: The energy team must account for the savings—or lack of savings–to senior management, using a feedback process so goals and targets can be reset if they aren't met. Use this process to verify energy use and to help identify any new opportunities for savings, and then to fine-tune your goals. If you fall short of your targets, you need to figure out why. Or perhaps you did meet your targets, but parts of the data were wrong. Perhaps the metering is off, or the energy accounting system needs readjustment. Regularly scheduled feedback from these reports will keep the energy master planning process up-to-date and realistic and will ensure accountability. One approach that has been successfully used in many organizations is the use of "energy report cards" and "intra-organizational listings/scorecards". Produced by the energy manager and sent out to all facilities/operating groups/departments on a monthly basis, these tools inform as well as motivate, based on the certainty that no one wants to be on the bottom of the list.

Communicate Results

Successful implementation of all stages of the EMP can be a useful vehicle for departmental—and personal—recognition. With a representative of the corporate PR office already on the energy team—enlist that individual's expertise for internal and external communications. Inform the organization's Board about the financial savings and improved asset value. Let the staff and community know about the environmental benefits. Apply for energy awards for national recognition in your sector. And use the good will generated by the success to keep the energy master planning process moving forward.

TIPS FOR SUCCESS

From these steps to an EMP, two points emerge as the most critical to success.

The first is ensuring buy-in from the top, with a long-term commitment and binding, formal statement of energy policy for the organization. Commitment to action also means that senior executives support the people in the middle—delegating authority to the energy manager or facility manager. If the EMP lacks commitment from the top down, no amount of effort by the energy manager will make it succeed. Once you've caught executives' attention, keep energy master planning on their radar screen by having the energy team build it into the organization's business plan. Such a commitment must also be apparent. Top management must continue to be seen (by all levels of staff) reaffirming their commitment, lest the EMP be perceived as just another short-term initiative of the organization.

The second critical component is line-management accountability, making specific individuals accountable for sustaining the savings. On the management side, executives need to mandate quantifiable goals and targets, strengthened by obligatory deadlines. Such requirements should even be built into job descriptions and evaluated as part of annual performance standards reviews. Likewise, incentives can be offered through these same personnel standards for individuals who meet their energy goals. Employee teams with day-to-day knowledge of the facility's operation can also be organized to identify additional opportunities for savings. At the upper level of the organization, corporate accountability for successful energy management can be communicated (and made public) through such mechanisms as the Global Reporting Initiative, especially in firms with sustainability principles that follow a "triple bottom-line".

CONCLUSION

Energy master planning is an effective and long-term shift in organizational cultures to address and adapt to the impact energy resources can have on their competitive posture and economic success. Energy is a universal raw material and is essential to the operation of almost every commercial, institutional, governmental, and even non-profit organization. Significant changes to the basic structure of the energy supply chain and significant energy price volatility, driven by a wide range of political and physical events (e.g., climate change, weather), have made energy both a competitive opportunity and a risk requiring active long-term planning to manage. Energy master planning requires buy-in from top levels of an organization down through the rank and file, selection/hiring of a dedicated energy manager, and creation of an energy team with membership from across the organization. This team must be given the resources and top-level access to develop and implement long-term planning for procurement and operations throughout the organization that places a premium on lasting energy-use reductions and cost optimization, rather than "burst" efforts that provide quick and quickly forgotten energy management efforts.

Successful organizations do not exist on a quarterly basis. They plan for the long-term in all aspects of their operations. Experience has shown that organization-wide adoption of EMP is effective in optimizing energy costs and needs to ensure competitive posture and success in the long-term business reality in which organizations exist. This is the new bottom line for energy.

ACKNOWLEDGMENTS

Permission has kindly been granted by the Energy Master Planning Institute for use of materials on its Web site and for materials published and presented by Coriolana Simon, founder of the EMPI.

REFERENCES

1. Allen, P.J.; Kivler, W.B. *Walt Disney World's Utility Efficiency Awards and Environmental Circles of Excellence*, Proceedings of WEEC 1995; Association of Energy Engineers: Atlanta, GA, 1995.
2. Tripp, D.E.; Dixon, S. *Making Energy Management "Business as Usual": Identifying and Responding to the Organizational Barriers*, Proceedings of the 2003 World Energy Engineering Congress; Association of Energy Engineers: Atlanta, GA, 2003.
3. Key, G.T.; Benson, K.E. *Collins and Aikman Floorcoverings: A Strategic Approach to Energy Management*, Proceedings of the 2003 World Energy Engineering Congress; Association of Energy Engineers: Atlanta, GA, 2003.

Energy Project Management

Lorrie B. Tietze
Interface Consulting, LLC, Castle Rock, Colorado, U.S.A.

Sandra B. McCardell
Current-C Energy Systems, Inc., Mills, Wyoming, U.S.A.

Abstract

Today's energy projects are becoming ever more complicated technically, beginning to involve more people, and beginning to include more integrated components and systems. As the complexity of an individual project increases, so does the importance of project management. Even if the responsibilities for project management are outside the business organization, good project management principles apply. To successfully manage an energy project, the project manager must have a plan by which he or she can direct project activities; the tools to document agreements and project work; the ability to facilitate communication; and the skills to manage the project's technical side (cost and schedule) as well as the project's human side (conflict resolution, team building, and gaining commitment from stakeholders). The process of energy project management is discussed here along with several tools that can assist managers in developing their own energy projects.

INTRODUCTION

Energy project management was once simple—meeting costs and schedules were considered sufficient. That is no longer true, and it is imperative that good project management principles are used and that the energy business owner remains intimately involved. The goal of this article, written for the energy business owner and the energy business project manager, is to discuss key requirements for successfully managing an energy project.

Today's energy projects are technically complicated, involve many people and disciplines, and consist of numerous integrated components and systems. In an energy project, the up-front sales and marketing, initial energy data collection, and data analysis take a tremendous amount of time and effort. After those efforts, there is often an assumption that project implementation will take care of itself. Of course that is not the case, and a poorly managed project will not deliver the intended results.

There are also time and cost advantages to good energy project management. Buildings last for 30–50 years and errors and omissions grow in importance and cost over time.

It is also less expensive to pay attention to project management than to ignore it. This allows potential problems to be identified and solved earlier and at a lower cost. The general rule is that if an item costs $1 to correct while the project is just an idea, it costs $10 during project definition, $100 during project design, $10,000 during project start-up, and $100,000 after the project is up and running.

To successfully manage an energy project, an energy project manager needs a plan by which he or she can direct project activities; the tools needed to document the agreements, the actual project work, and to facilitate communication; and the skills to manage the technical side (cost and schedule) and the human side (conflict resolution, gaining commitment) of the project. This article will introduce basic energy project definitions, an energy project management plan, and the tools and skills necessary for the successful energy project.

DEFINITIONS

Commissioning. The systematic process of assuring by verification and documentation—from the design stage to a minimum of one year after construction—that all building facility systems perform interactively in accordance with the design documentation and intent and with the owner's operational needs, including the preparation of operation personnel.

Energy project management plan. A set of organized activities, each having a definite beginning and end, that when completed by a project team, deliver a specific business objective.

Milestone. Key, high-level summary activities that must be completed by a set date.

Operation. Business operation for which the energy project was performed.

Keywords: Energy project management; Commissioning; Start-up; Project stages; Performance bid; Energy efficiency; Schedule risk analysis.

Energy Project Stage	Stage Deliverables
Conceptual Stage	▪ Project's general description or project brief, ▪ Needs basis, ▪ Return on Investment (ROI) estimate, such as a Life Cycle Cost Analysis (LCCA), ▪ Initial cost range, ▪ Initial benefit range (including intangibles, where appropriate), ▪ Initial project timing, ▪ Project Charter, ▪ Initial project management team.
Definition Stage	▪ Detailed project scope, ▪ Detailed project cost, ▪ Initial project funding strategy, ▪ Bid process strategy, ▪ Updated Return on Investment (ROI) or Life Cycle Cost Analysis (LCCA), ▪ Project Management strategy, ▪ Final Project management team, ▪ Initial project schedule, ▪ Initial delivery plan, ▪ Initial skilled resource personnel plan, ▪ Initial communication plan, ▪ Initial operational and maintenance training plan, ▪ Major risk and challenges, ▪ Initial contingency plans.
Design Stage	▪ Detailed engineering designs, ▪ Detailed project plan and schedule, ▪ Detailed skilled personnel resource plan, ▪ Risk analysis and contingency plan, ▪ Financing plan, ▪ Operations and maintenance plan, ▪ Commissioning plan, ▪ Operations plan, ▪ Building modeling as appropriate,

Fig. 1. Energy project stages and deliverables.
Source: From Interface Consulting, LLC 2005.

	▪ Sustainability analysis and design plan, ▪ Bid for funding.
Construction Stage	▪ Detailed construction schedule, ▪ Detailed construction personnel resource plan, ▪ Coordination plan for sub-contractors providing different system components, ▪ Construction communication plan, ▪ Initial start-up and ongoing operations plan, ▪ Updated operational and maintenance training plan.
Delivery Stage	▪ Detailed start-up schedule, ▪ Detailed operational and maintenance personnel resource plan, ▪ Detailed start-up and ongoing operations plan, ▪ Detailed turnover plan, ▪ Updated operational and maintenance training plan, ▪ Project Post mortem.

Fig. 1 (*Continued*)

Project stage. A logical sequence of activities, milestones, and deliverables.

Schedule risk analysis. Equipment and computer programs that let users measure, monitor, and quantify schedule impact on their energy project and help identify schedule risks and opportunities.

ENERGY PROJECT MANAGEMENT PLAN

The energy project management plan is the roadmap for the project. It is a set of organized activities, each having a definite beginning and end, intended to meet a specific business objective when completed by a project team. To make the management of the process easier, project activities are divided into separate structured stages.

Each project stage contains a logical sequence of activities, milestones, and deliverables required to achieve the project objectives. The milestones are key, high-level summary activities that must be completed by a set date, and are linked to the success criteria for each particular stage. The number of stages and milestones is dependent upon the energy project plan length and complexity. The deliverables ensure that all prerequisite work is completed prior to moving to the next stage of the project. Transitions from one stage to the next are times to assess and review both the stage just completed and the project plan.

Fig. 1 illustrates the typical stages and deliverables in an energy project plan.

The plan's main purpose is to provide the roadmap for specific project activities, but it is also a collaboration, alignment, and communication tool. It details which activities are happening at what time, who is involved, and what is happening next. It is also an acknowledgement tool—recognizing when the project team has achieved milestones and when it is on track to delivering the overall project objectives. When the energy project plan is well crafted, the plan followed, and the success criteria realized, an energy project manager can produce maximum results in delivering the overall project objectives.

Conceptual Stage

The conceptual stage is the beginning of an energy project. Its deliverables are broad and are intended to evaluate the project's appropriateness for the business and its needs by answering two key questions:

- "Is this project worth a level of investment?"
- "How does this project further our business plan and mission?"

Business stakeholders answer these questions and develop the initial cost estimate. At this point, there is uncertainty in the project scope and associated costs. While strategic level decisions can be made with these numbers, the conceptual stage costs should not be used as the project cost commitment.

If the project is a fit for the business, a project summary or project brief is written and then reviewed with key stakeholders. The project brief reviews have two purposes—to educate and to gain alignment and commitment to the proposed project. Once the project has commitment from all appropriate parties, a project charter is created. The project charter includes much of the same information as the project brief, as well as the names of the project team and project authorization signatures. It is the official document authorizing the project manager to begin work. Once the charter is approved, the project moves into the first tactical stage—definition.

Definition Stage

This is the stage at which the "how?" and the "when?" for an energy project are defined. "Definition" requires the largest amount of documentation and the most deliverables of any stage and it builds on information from the conceptual stage. Project work and scope is identified for each step in the project.

The scope detail includes electrical and mechanical layout, as well as civil and structural where appropriate, and the general layout of major pieces of equipment, piping, and electrical systems. A physical inspection of the planned location is recommended so that obstacles can be identified as early as possible. For LEED certification (U.S. Green Building Council's Leadership in Energy and Environmental Design), as well as for good business practice, all interactions between building systems should also be taken into consideration from this point forward. A third party "owner's agent" or "commissioning agent" is also a valuable team member who can ensure that the building is both designed and constructed in an energy efficient way and that the finished product meets the project objectives.

The initial personnel resource plan for the project should be completed as early as possible, including a decision on whether to have the project managed by in-house personnel or to use project management services provided by an outside vendor. Remaining personnel are identified iteratively during the scope refinement and activity identification stage and include core project team members, technical resource members, support members, operating department members, and outside vendor or resource team members. All personnel, titles, and general responsibilities should be summarized in a single contact document for wide distribution throughout the project and the business.

Once the project personnel are named and engaged, a communication plan is developed. The plan specifies how project information will be collected, what information will be disseminated, who the information will go to, and when the information should be communicated. A RACI chart (discussed in "Collaboration and Alignment Tools") provides structure to this communication plan throughout the project.

All project activities are planned for during the definition stage and it is critical to remember that the construction and training activities inherent in energy projects involve large numbers of people—so the human elements of the project must be considered at all times. Due diligence in definition provides a basis for all downstream work and provides important details for communication with all personnel involved in the project. This is the first stage where all people involved will have enough detail to see what the project will really look like, how it will impact them, and what steps should be taken to minimize the impact on existing business operations.

Another face of the critical human element in energy projects that carries heavy operational impact is training of the operational and maintenance personnel. Training should be planned during the definition stage and is developed by the following model:

- Step 1—Identify the new tasks and activities.
- Step 2—Identify the skills to support the Step 1 tasks.
- Step 3—Identify the knowledge and demonstration to show mastery of the Step 2 skills.
- Step 4—Identify the training required to accomplish Step 3.

Once the steps are identified, corresponding theoretical and practical project activities are incorporated into the project schedule.

Another key milestone document in this stage is the delivery plan, which includes success criteria for the project completion substages of start-up, full-going, and turnover. The success criteria are negotiated and mutually agreed to by both the project team and the business operation. The delivery plan documents what the project team is promising to provide for the operation and what the operation promises to provide for the project. When designed well, the delivery plan is a powerful communication tool that can engender agreement among all partners throughout the life of the project.

Start-up criteria are designed to demonstrate that equipment and processes can function at the engineered design rates, quality, and efficiency, usually for a normal business operating shift. Start-up is led and mainly staffed by the energy project team members with operations personnel assisting.

The full-going criteria are designed to demonstrate that the equipment, processes, and people are capable of running at the engineering design rates, efficiencies, and specified quality for 2–7 days. Full-going is led by either project team or operational personnel, but staffed by operational personnel.

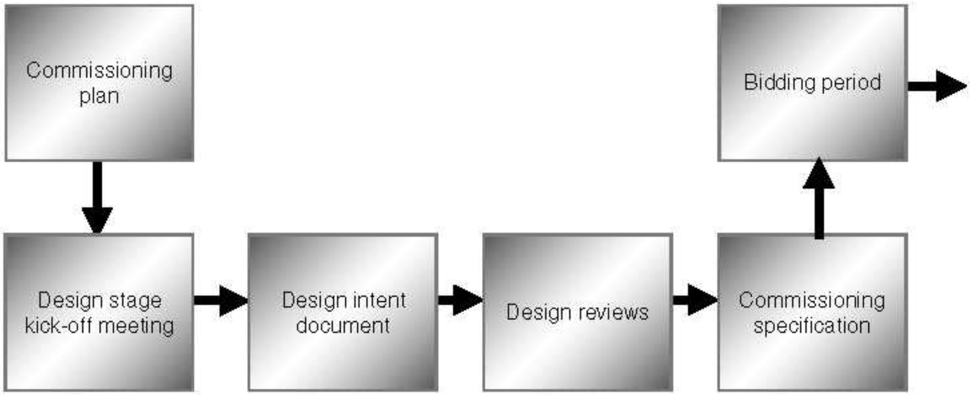


Fig. 2 Design stage commissioning activities.
Source: From Sebesta Blomberg and Associates, Inc., and Current-C Energy Systems, Inc. (see Ref. 1).

Turnover criteria additionally include the personnel performance criteria (operational and maintenance), training criteria (operational and maintenance), documentation criteria, and the resolution of any outstanding items such as study or research items. Once all of the turnover criteria are achieved, the system is turned over to the operational personnel and the project personnel's work is nearly complete; it remains only to complete any outstanding follow-up activities and conduct a project postmortem.

Also during the definition stage, a baseline project schedule should be developed with broad participation by the project team, operational personnel, outside resources, logistics resources, etc. The process used is to list the project stages and individual activities in each stage until all of the project activities are identified. The activities should then be linked to one another by identifying each activity's predecessors and successors and finally, durations should be assigned to each activity using 3-pt duration estimates (minimum, probable, maximum) to support the use of schedule risk analysis (see Ref. 2). This critical step yields an assessment of the statistical probability of completing the project by a certain time and is extremely helpful when analyzing the impact of scope or resource changes on a project.

The final steps in the definition stage are to update the cost estimate, finalize the funding strategy and methodology for the project, and determine the preferred bid structure.

Design Stage

The design stage goal involves development of designs for all equipment, processes, and systems. This stage frequently involves outside resources with expertise in detailed engineering design and often requires additional study or research in an effort to predict the operation and interaction of the full energy project on a small scale. Techniques such as pilot projects and system modeling may provide critical information as well as initial training for both project and operational and maintenance personnel and can also gain valuable input and commitment into the project. When engineers, who design the equipment and systems and know how they should operate, work with operational and maintenance personnel, who know how the equipment will really be operated, the results are sometimes surprising.

For energy projects, the interplay between initial cost and long-term expenses (utility, maintenance, and other) is also a critical part of the design stage. For example, a component with a 1-year life costing \$1000 may be upgraded to a component with a 5-year life that costs \$2000.

Energy system commissioning is the next step in design—it is the systematic process of assuring by verification and documentation from the design stage to a minimum of one year after construction that all building facility systems perform interactively in accordance with the design documentation and intent and with the owner's operational needs. Commissioning ensures that the energy project gives the business the promised results from the project including documentation, testing, and training. Specific commissioning activities are included as appropriate in the start-up, full-going, and turnover plans. Integrated design stage commissioning activities as practiced by companies such as Sebesta Blomberg are shown in Fig. 2.

Bid Preparation

It should be noted that bid preparation is an ongoing task that is parallel to the definition phase. It is treated separately here because of its complexity and importance.

The main types of bids in energy projects are design/build, fixed bid, not-to-exceed, performance bids, and hybrids. The appropriate type of bid depends upon such criteria as the type of energy project, the level of design detail, technological complications, the schedule, funding available, and complexity of the project.

Design/build bids are the best choice when scope is not well defined and there are many unknowns—either in scope or personnel experience. The disadvantage is that

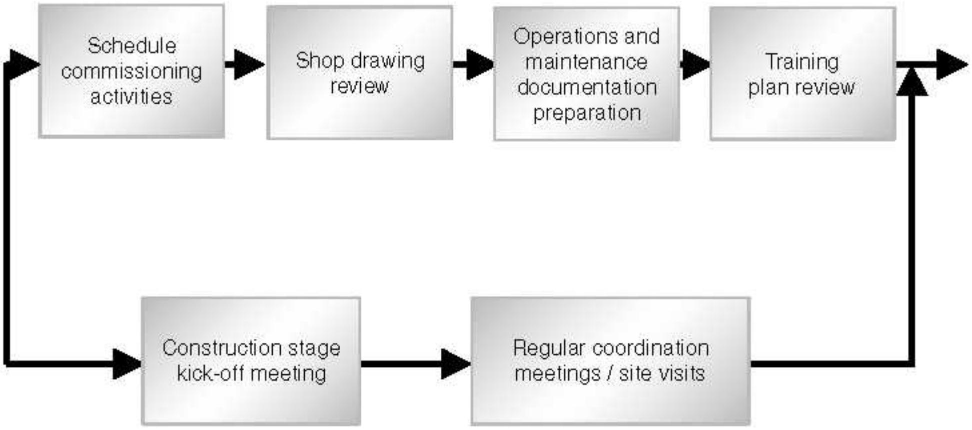


Fig. 3 Construction stage commissioning activities.
Source: From Sebesta Blomberg and Associates, Inc., and Current-C Energy Systems, Inc. (see Ref. 1).

design/build costs grow proportionally as the scope changes, with the costs passed through to the project. Consequently, the design/build bid can be the highest priced, but it provides greater flexibility to handle the issues or scope changes during the project.

Fixed bid is the best choice when the scope is well defined. There are few unknowns and project personnel are experienced enough to strongly defend and enforce the agreed upon scope of work and remove excessive contingency fees from the bid price. Scope additions or changes not requested by the operation result in those additional costs coming out of the bidder's pocket. Consequently, the fixed bid can be the lowest priced option while also providing the least flexibility.

The not-to-exceed bid is a hybrid. A project with a well-defined scope uses the fixed bid methodology and one with a less defined scope uses the design/build methodology. There is an added line item for the estimated cost of scope changes. A not-to-exceed bid is usually higher than a fixed bid, but lower than the eventual design/build bid.

Performance bid (or performance contract, "PC") is also a hybrid and combines any of the above methodologies with energy efficiency performance, which gives the supplier an incentive to ensure that the project meets the promised business objectives. In the energy project industry, an endless variety of different techniques have evolved for particular situations or needs, but they share the characteristic that the project is paid for partly based upon its adherence to prespecified performance criteria, and the contractor covers part of the risk for the project performing as promised.

Finally, the bid can also be affected by regulatory compliance and energy efficiency standards or regulations. New energy project equipment must meet the current regulatory requirements and existing equipment may have to be upgraded if it will be linked to the new equipment. ASHRAE standards for energy efficiency are used as standards in some jurisdictions, and as the U.S. Green Building Council's LEED certification and EnergyStar plaques for "high performance" or "green" buildings become more popular, those standards, processes, and procedures should also be included in the bid process.

Construction Stage

Traditionally this is when 'project management' starts, although as we have seen that it is really the fifth stage in the process. The goal of the construction stage is to build and install the equipment and processes for the project and the deliverables include schedules, resource plans, and refinements and adjustments to the communication and delivery plans.

For the detailed construction plan, the operator must decide whether to manage the construction in-house or to employ a general contractor for construction management and then choose the format for the construction schedule (number of days/week and hours/day). Once these are decided, the construction personnel needs are broken down by specific trade skills; union and nonunion workers; roles and responsibilities of core construction personnel, sub-contractors, and operational resources; and a schedule of the resources is produced.

The communication plan is updated with a plan for communicating daily and weekly construction activities and other information, including a feedback loop for rapid adjustments to construction schedule changes. Energy system commissioning activities continue, and are shown in Fig. 3.

Delivery Stage

The delivery stage can be the most chaotic stage as it includes the initial start-up, training, and turnover of the equipment and processes to the operation. The goal is to get everything started up as designed (rate, efficiency, and quality), provide training for all operational and maintenance personnel, turn over the equipment and processes, and close out the project. In a manner similar to the construction

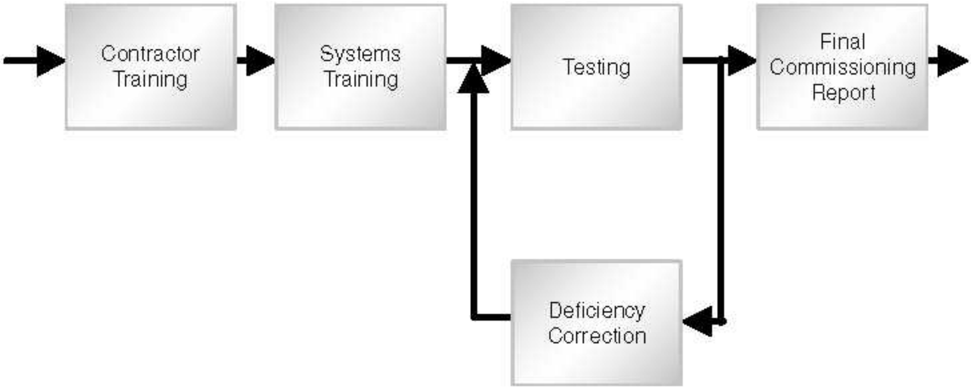


Fig. 4 Delivery stage commissioning activities.
Source: From Sebesta Blomberg and Associates, Inc., and Current-C Energy Systems, Inc. (see Ref. 1).

stage, the project delivery personnel needs are broken down by specific trade skills, roles and responsibilities, and a schedule of when each set of the personnel resources will be needed. The personnel requirements at this point are very large, especially from the ongoing operation.

The project schedule will have previously identified training activities, personnel, and schedules; start-up, full-going, and turnover activities personnel and schedules; and delivery stage commissioning activities (illustrated in Fig. 4). During the project delivery stage, those plans are followed as they have been adjusted during the previous project management stages, and the project is then completed.

Upon completion of the delivery activities, the achievement of the delivery criteria, and the turnover of the project to the operation, the project team conducts a postmortem on the project. The purpose of the postmortem is to celebrate the successes, identify the lessons learned, and to document both of these so that the next project has an even greater chance of success.

The commissioning activities extend—with the operational personnel—past the closure of the main project during the warranty period for the equipment. The postproject commissioning activities are illustrated in Fig. 5.

TOOLS FOR ENERGY PROJECT MANAGEMENT

Tools can make it easier and faster to execute project tasks, provide a standard methodology, enhance communication, and minimize confusion. The following are examples of typical tools used during energy projects.

Scope Tools

The first set of tools help manage energy project scope and provide documentation for both the equipment included in the project and for the activities that the project team will perform during the project. They include:

- Project brief.
- Project charter.
- Start-up, full-going, turnover.
- Scope change management.

The project charter, introduced in "Conceptual Stage", defines the project and provides a means to secure both understanding and commitment from all stakeholders. A sample format is shown in Fig. 6.

As the project proceeds, the start-up, full-going, turnover tool is the most important communication tool between the formal project team and the business operation. It documents the agreements between the two organizations regarding equipment performance, system performance, training completion, etc. and it is the basis for declaring portions of the project complete and for releasing project personnel. This tool must be created specifically for each project, but an example is shown in Fig. 7. For each process, the tool lists the project equipment and processes and determining performance criteria for the start-up, full-going, and turnover.

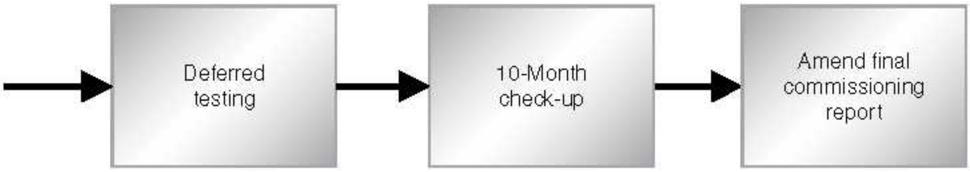


Fig. 5 Warranty period commissioning testing.
Source: From Sebesta Blomberg and Associates, Inc., and Current-C Energy Systems, Inc. (see Ref. 1).

PROJECT CHARTER	
Date:	
Project:	
Project objective:	
Business basis for project:	
Initial cost:	
Initial return:	
Initial timing:	
Major risks:	
Project sponsor(s):	
Project team:	
Project authorization:	

Fig. 6 Project charter.
Source: From Interface Consulting, LLC, 2005.

Scope management is an ongoing challenge as scope change requests come from all personnel involved in the project. Criteria should be established to sort out small, inconsequential change requests from larger ones having greater impact and a standardized scope change tool should be developed for approval and documentation. For larger change requests, a formal review should evaluate the change's impact on the project's cost, schedule, and other related systems. Its impact on the schedule can then be quantitatively evaluated using a schedule risk analysis.

START-UP, FULL-GOING, TURNOVER TOOL			
	Start-up criteria	Full-going criteria	Turnover criteria
Process #1- includes Equipment #1 and Equipment #2	Run 8 continuous hours at 80% efficiency, producing 90% quality	Run 7 continuous days, 24 hours/day at 85% efficiency, producing 95% quality	- Start-Up and Full-Going criteria are met. - Operational and maintenance personnel trained and qualified on required skills. - Equipment documentation and files established and in the operation including (drawings, manuals, spare parts lists, etc.).

Fig. 7 An example of the start-up, full-going, turnover tool.
Source: From Interface Consulting, LLC 2005.

Scheduling Tools

The scheduling tools for energy projects are generally computerized and include two basic types—project scheduling software and schedule risk analysis software. Project scheduling programs project the project delivery date by linking together individual activity durations and dates from the project team to create a critical path schedule. With such a program, the impact on the schedule end date caused by interim changes becomes very clear.

After a project schedule has been created, it can be analyzed for schedule risk. Even though traditional critical path schedules are designed to show the most important activities in the project schedule, they do not take uncertainty into account. Consequently, if activities do not occur as predicted, the critical path may change, and with it the focus of the project team (see Ref. 2).

Schedule risk analysis uses 3-pt duration estimates for each activity (minimum, probable, and maximum), creating a distribution of durations for each activity. The analysis is conducted by using statistical techniques to create a new distribution of project end dates and probabilities, quantifying the probability of completing the project on time.

More importantly, for each simulation, the critical path, the critical path activities, and the activities that most effect the project end date are noted and summarized as they are calculated. From schedule risk analysis, they remain the important activities even as the project schedule changes. The project manager therefore has fewer key activities to focus on for the successful delivery of the project.

Collaboration and Alignment Tools

One of the challenges during an energy project where systems and variables are interactive is the volume of information and the number of decisions handled each and every day. No individual can be involved in every aspect of the project, and a tool such as RACI helps prevent gaps and overlaps in project team activities. RACI stands for:

- Responsible—who owns the project or the problem.
- Accountable—who must sign-off or approve the activity, and who the responsible person reports to.
- Consult—who provides input but isn't responsible for the activity.
- Inform—who needs be told but not consulted about the activity.

To use the RACI tool, identify the areas of decisions or tasks/activities involved in the energy project, identify the roles involved in the project, and complete the tool by filling in the Rs, As, Cs, and Is. Ideally, for any individual activity, only one person should have the 'R.' As the tool is

	Project sponsor	Project manager	Operation resource	Technical resource	Support resource
Project activity #1	A	R	I	C	I
Project activity #2	R	C	C	C	I
Project activity #3	C	C	R	I	C

Fig. 8 RACI chart.
Source: From Legacy business records project-library and archives Canada, 2005.

completed for each activity, task gaps and overlaps become quickly visible—allowing the project team to resolve them (Fig. 8).

SKILLS FOR ENERGY PROJECT MANAGEMENT

The energy project manager needs both technical project management skills and people project management skills to deliver a project on time, on budget, and with excellence.

Technical Project Management Skills

The technical project management skills have a long history and are well understood, and they include:

- Scope definition.
- Scope management.
- Scheduling.
- Cost management.

These skills are critical to managing the nuts and bolts of the project. Training for the technical project management skills is wide-spread and readily available from such organizations as the Association of Energy Engineers (AEE) and others. For this reason, technical skills are not addressed in this paper.

Personnel Management Skills

As project circumstances change, people skills can add a new layer of effectiveness for a project manager. For energy efficiency projects where multiple technical disciplines are included as core parts of the team, people skills are even more critical. Energy project managers must be skilled at choosing their project team, gaining commitment to the project, and resolving conflict (see Ref. 3).

The first critical decision is the composition of the project team. Each individual's qualifications, skills, attitude, and commitment should all be considered. For example, a person who has great technical skills but who treats team members poorly should be considered only after careful thought as to the amount of time the project manager is willing to spend listening to complaints and refereeing between team members, and contingency plans should be developed to minimize problems. Unresolved people issues tie up personnel and their productivity, slowing down both the quality and the completion of the project.

The energy project manager should also evaluate the team's communication and people skills. A good project team can produce trust and handle confrontation. It can almost be said that the team becomes the human representation of the project and it must be trained and prepared to handle people responsibilities as competently as technical responsibilities.

One key for energy project managers is gaining the commitment and cooperation of the operational personnel. Gaining commitment to a project is a process, not an initial announcement followed by informational updates. People go through a predictable process in becoming committed to something and the faster the project manager can recognize the process, the faster they can foster commitment.

When a new energy project is first proposed, the stakeholders, project team, and operational people all have an emotional reaction, although the reactions vary and change as they find out more about the project and its impact. If the project provides them with tangible benefits, they become committed very quickly. For many people, however, the stated benefit is a positive business benefit with no obvious personal impact. Consequently, the typical first reactions are either indifference or anger, and the energy manager should never confuse indifference with commitment.

Indifference is characterized by a "wait and see" attitude and when comments like "I've seen all of this before and I'll see it again" or "Hmm—we'll just see if they really can do what they say they can do" are made. As people progress through indifference, they often become angry and focus that anger on the energy project manager who is the visible symbol of the project. Complaints fly and the project is described as "terrible". The project manager must at this point remain positive and committed, which can be a challenge as it often seems few people are on his or her side. The project manager can even be tempted to avoid conversations with those who are angry, hoping that avoidance and time will smooth away their anger. Unfortunately, this is usually counterproductive, as motives are created to explain the project manager's avoidance and the situation spirals down.

Paradoxically, it is at this point that project managers have an opportunity to listen and learn. At the time, listening is likely to seem like the long, slow way to move the project forward but, once people feel they have been heard, they often move from saying "the entire project is awful", to "the project won't work because…". The "because" is very important, and it is important that the project manager hear and recognize this change in order to actively involve people and bring them the rest of the way to commitment. Once they agree to help be part of solving the "because", they are committed to the project. Energy projects bring change, and change always causes conflict. The project manager must be the first one to pull out a hidden conflict and set it squarely on the table for resolution. He or she must be completely intolerant of arguments vs. resolution and expect the same from the entire team.

CONCLUSION

Today's energy project manager needs to understand the plan, the tools, and the skills necessary to successfully manage the complete energy project—from costs and schedules to conflict resolution and commitment. A well-defined project management plan sets out a path that defines, creates, measures, and documents success. Using project management tools makes the process easier and leaves a trail of documentation regarding project decisions. A project manager with both technical and human skills is able to effectively manage both scope and people issues, saving time and money. Whether the project is managed in-house or a contract energy project manager is used, these three areas provide the energy project manager with what is needed to successfully deliver an energy project.

REFERENCES

1. Heneghan, T.; Sandra, M., *Commissioning Activities for Energy Project Stages*, Sebesta Blomberg and Associates, Inc., 2005.
2. Tietze, L.; *Why Schedule Risk Analysis?*, True Project, LLC, Fall/Winter, 2005.
3. Tietze, L.; *Integrating the Human Side of a Technology*, Interface Consulting LLC, 2004.

Energy Service Companies: Europe

Silvia Rezessy
Energy Efficiency Advisory, REEEP International Secretariat, Vienna International Centre, Austria, Environmental Sciences and Policy Department, Central European University, Nador, Hungary

Paolo Bertoldi
European Commission, Directorate General JRC, Ispra (VA), Italy

Abstract
Energy service companies (ESCOs) are important agents to promote energy efficiency improvements. This entry attempts to address a major gap identified in the process of conducting the first in-depth survey of ESCO businesses in Europe: the lack of common understanding and terminology related to types of energy service providers, contractual and financing terms, and the implications of different models. It provides the key terminology related to energy services, ESCOs, contractual models, and financing structure.

INTRODUCTION

Energy service companies (ESCOs) are important agents to promote energy-efficiency improvements. The initial ESCO concept started in Europe more than 100 years ago and then moved to North America. In the past decade, Europe has seen an increased interest in the provision of energy services that has been driven by electricity and gas restructuring and the push to bring sustainability to the energy sector. The actual market for energy-efficiency services in Western Europe was estimated to be 150 million euro per annum in 2000, while the market potential for Western Europe was estimated to be 5–10 billion euro per annum (these estimates are according to our definitions of ESCO, energy performance contracting (EPC), and third-party contracting (TPF) (see definitions later), and do not include energy supply) (Bertoldi et al. 2003 and references herein).[2,14] A survey of the U.S. ESCO industry estimates ESCO industry project investment reaching $1.8–$2.1 billion U.S. in 2000, with industry revenue growing at 9% per year in the period 1996–2000—down from almost 25% in the previous 5 years.[6]

This entry attempts to address a major gap identified in the process of conducting the first in-depth survey of ESCO businesses in Europe: the lack of common understanding and terminology related to types of energy service providers, contractual and financing terms, and the implications of different models. It provides the key terminology related to energy services, ESCOs, contractual models, and financing structure, pointing at the implications of different choices.

This entry uses terminology agreed upon after a long process of consultation undertaken in 2004 (the entire report is available for free download at http://energyefficiency.jrc.cec.eu.int/pdf/ESCO%20report%20final%20revised%20v2.pdf) and early 2005 with major stakeholders in Europe and the United States. The entry builds on the status report "ESCOs in Europe," published in 2005 by Directorate General Joint Research Center of the European Commission. The report has covered the 25 member states of the European Union (EU), the New Accession countries Bulgaria and Romania, Switzerland, and Norway.[1]

ENERGY SERVICES, ENERGY SERVICE COMPANIES, AND PROJECT ELEMENTS

Energy services include a wide range of activities, such as energy analysis and audits, energy management, project design and implementation, maintenance and operation, the monitoring and evaluation of savings, property/facility management, energy and equipment supply, and provision of service (space heating/cooling, lighting, etc.).

Consultant engineering companies specialized in efficiency improvements, equipment manufacturers, energy suppliers or utilities, which provide energy services for a fixed fee to final energy users or as added value to the supply of equipment or energy, are referred to as *Energy Service Provider Companies (ESPCs)* (note that this is a very different use of the acronym ESPC common in North America, where it means energy service performance contracting). Energy service provider companies may have some incentives to reduce consumption, but these are not as clear as in the ESCO approach (see explanations below). Often, the full cost of energy services is recovered in the fee, so the ESPC does not assume any risk in case of underperformance. Energy service provider companies are

Keywords: Energy service companies; Energy performance contracting; Financing models.

paid a fee for their advice/service rather than being paid based on the results of their recommendations.[13] Principally, projects implemented by ESPCs are related to primary energy conversion equipment (boilers, combined heat and powers [CHPs]). In such projects, the ESPC is unlikely to guarantee a reduction in the delivered energy consumption because it may have no control of or ongoing responsibility for the efficiency of secondary conversion equipment (such as radiators, motors, and drives) and no control of the demand for final energy services (such as space heating, motive power, and light).[12]

Energy service companies also offer these same services. Energy service companies are fundamentally different from ESPCs, and ESCOs' activities can be distinguished from ESPCs' activities in the following ways:

- Energy service companies in Europe guarantee the energy savings and the provision of the same level of energy service at a lower cost by implementing an energy-efficiency project (in North America, ESCOs do not necessarily guarantee energy savings. If they do, the document that specifies this guarantee is called an energy savings performance contract). A performance guarantee can take several forms. It can revolve around the actual flow of energy savings from a project; it can stipulate that the energy savings will be sufficient to repay monthly debt service costs for an efficiency project; or it can stipulate that the same level of energy service will be provided for less money.
- The remuneration of ESCOs is tied directly to the energy savings achieved.
- Energy service companies typically finance (or assist in arranging financing) the installation of an energy project that they implement by providing a savings guarantee.
- Energy service companies may retain an ongoing operational role in measuring and verifying the savings over the financing term.

Typically, a project developed and implemented by an ESCO includes the following elements/steps:

- Site survey and preliminary evaluation
- Investment-grade energy audit
- Identification of possible energy-saving and efficiency-improving actions
- Financial presentation and client decision
- Guarantee of the results by proper contract clauses
- Project financing
- Comprehensive engineering and project design and specifications
- Procurement and installation of equipment; final design and construction
- Project management, commissioning, and acceptance
- Facility and equipment operation and maintenance for the contract period
- Purchase of fuel and electricity (to provide heat, comfort, light, etc.)
- Measurement and verifications (M&V) of the savings results.

The investment-grade audit (IGA) deserves special attention. The traditional energy audit does not sufficiently consider how implemented measures will behave over time. Because auditors must consider the conditions under which measures will function during the life of the project, an IGA builds on the conventional energy audit. Unlike the traditional energy audit, which assumes that all conditions (related to system, payback, and people) remain the same over time, an IGA attempts to predict a building's energy use more accurately by adding the dimension of a risk assessment component, which evaluates conditions in a specific building or process. Aspects of the IGA include risk management, the "people" factor, M&V, financing issues, report presentation guidelines, and master planning strategies.[9]

FINANCING OPTIONS

Three broad options for financing energy-efficiency improvements can be distinguished. The approach to financing is just one factor that shapes the structure of an EPC. Other factors relevant for the type of financing arrangements and repayment structures include the allocation of risks, the services contracted, and the length of the contract.[11]

Energy service company financing refers to financing with internal funds of the ESCO; it may involve use of its own capital or funding through debt or lease instruments. Energy service companies rarely use equity for financing, as this option limits their capability to implement projects on a sustainable basis.

Energy-user/customer financing usually involves financing with internal funds of the user/customer backed by an energy savings guarantee provided by the ESCO. For instance, a university can use its endowment fund to finance an energy project, in which the energy savings are guaranteed by an ESCO. Energy-user/customer financing may also be associated with borrowing in the case when the energy user/customer, as a direct borrower, has to provide a guarantee (collateral) to the finance institution. The provision of collateral by the consumer itself is what distinguishes customer financing from third-party financing (TPF).

Third-party financing refers mostly to debt financing. As its name suggests, project financing comes from a third party (e.g., a financial institution) and not from internal funds of the ESCO or of the customer. The finance institution may either assume the rights to the energy savings or take a security interest in the project equipment.[13] Two conceptually different TPF arrangements are associated with EPCs, and the key difference between them is which party borrows the money: the ESCO or the client.

- The first option is that the ESCO borrows the financial resources necessary for project implementation.
- The second option is that the energy-user/customer takes a loan from a financial institution, backed by an energy savings guarantee agreement with the ESCO. The purpose of the savings guarantee is to demonstrate to the bank that the project for which the customer borrows will generate a positive cash flow (i.e., that the savings achieved will certainly cover the debt repayment). Thus, the energy savings guarantee reduces the risk perception of the bank, which has implications for the interest rates at which financing is acquired. The cost of borrowing is strongly influenced by the size and credit history of the borrower.

When the ESCO is the borrower, the customer is safeguarded from financial risks related to the project's technical performance because the savings guarantee provided by the ESCO is either coming from the project value itself or is appearing on the balance sheet of the ESCO; hence, the debt resides on someone else's balance sheet (ESCO's or financial institution's). Both public and private customers can benefit from off-balance-sheet financing because the debt service is treated as an operational expense and not a capital obligation; therefore, debt ratings are not impacted. For highly leveraged companies, this is important, because the obligation not showing up on the balance sheet as debt means that company borrowing capacity is freed.[10] However, different countries apply various conditions that need to be met for financing to be viewed as an operating lease, for example. Unless those conditions are met, financing is automatically considered to be a capital lease. Therefore, parties seeking financing first need to inquire about the country-specific conditions for operational financing.

Large ESCOs with deep pockets (hence, high credit ratings) have started to prefer TPF to their own funds because their costs of equity financing and long-term financing are often much greater than what can be accessed in the financial markets. Also, if an ESCO arranges TPF, its own risk is smaller. This would allow for a lower cost of money for the same level of investment, so more money would be assigned to the project.[8] The cost associated with nonrecourse project financing by a third party (e.g., one in which project loans are secured only by the project's assets) is the highest, as it entails more risk and, hence, higher interest rates.[13]

Furthermore, as already mentioned, equity contributions from the ESCO are often deemed undesirable by ESCOs, as they tie up capital in a project. This emphasizes the fundamental concept that an ESCO is a service company and not a bank or a leasing company. The primary reason that ESCOs do not and should not provide project financing with internal funds is that it makes their balance sheets look like they are banks and not a service companies. Local practices, the inability of customers to meet financiers' creditworthiness criteria, and costs of equity financing are some of the factors that determine whether ESCOs will provide financing (debt or equity).[8] Small or undercapitalized ESCOs, which cannot borrow significant amounts of money from the financial markets, prefer their role not to be financing energy-efficiency investment.

ENERGY PERFORMANCE CONTRACTING MODELS

Energy performance contracting is a form of creative financing for capital improvement that allows funding energy-efficiency upgrades from cost reductions. Under an EPC arrangement, an external organization (ESCO) develops, implements, and finances (or arranges financing for) an energy-efficiency project or a renewable energy project and uses the stream of income from the cost savings or the renewable energy produced to repay the costs of the project, including the costs of the investment. Essentially, the ESCO will not recover all of its costs unless the project delivers all of the energy savings guaranteed. The approach is based on the transfer of technical risks from the client to the ESCO based on performance guarantees given by the ESCO. In EPC, ESCO remuneration is based on demonstrated performance; a measure of performance is the level of energy or cost savings, or the level of energy service. Therefore, EPC is a means to deliver infrastructure improvements to facilities that lack energy-engineering skills, manpower or management time, capital funding, an understanding of risk, or technology information. Cash-poor yet creditworthy customers, therefore, are good potential clients for EPC. Fig. 1 illustrates the EPC concept.

Energy performance contracting is usually distinguished from energy supply contracting (delivery contracting) that is focused on the supply of a set of energy services (e.g., heating, lighting, motive power) mainly via outsourcing the energy supply. Chauffage (see details later in this entry) includes supply contracting. In contrast, EPC typically targets savings in production and distribution.

Energy performance contracting is risk management and effective ESCOs have learned to use project financial structures to help manage the risks. Below we examine different contracting models and the conditions under which they deliver at their potential.

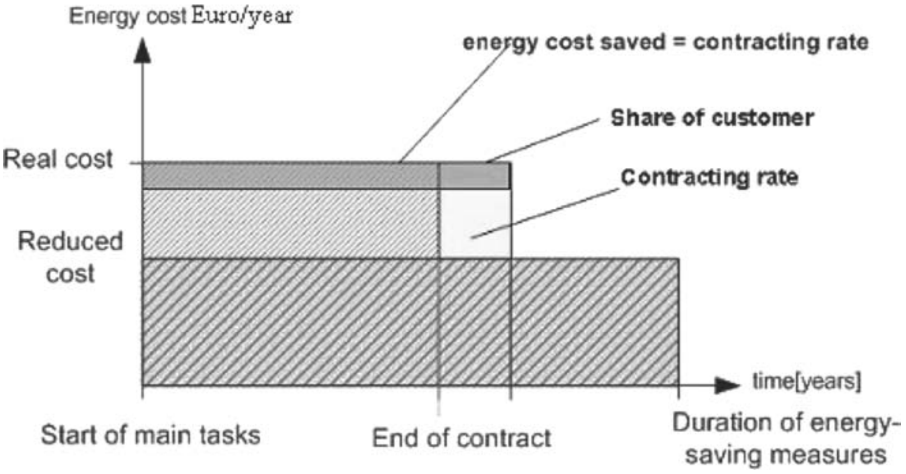


Fig. 1 Energy performance contract. *Note: Note that on this graph, real cost refers to the initial cost, while the contracting rate depicts the cost savings, which in this case are shared between the customer and the energy service company (ESCO) (see the explanation on shared savings). Source: From Berlin Energy Agency.

Guaranteed Savings and Shared Savings

Fig. 2 illustrates the relationships and risk allocations among the ESCO, customer, and lender in the two major performance contracting models: shared savings and guaranteed savings. Brief descriptions are also given. An important difference between guaranteed and shared savings models is that in the former case, the performance guarantee is the level of energy saved, whereas in the latter, this is the cost of energy saved.[7,11]

Under a guaranteed savings contract, the ESCO assumes the entire design, installation, and savings performance risks but does not assume credit risk of repayment by the customer. Consequently, guaranteed savings contracts are not usually applicable to ESCO financing provided internally or through TPF with ESCO borrowing. The projects are financed by the customers, who can also obtain financing from banks, from other financing agencies, or a TPF entity. The key advantage of this model is that it provides the lowest financing cost because it limits the risks of the financial institutions to their area of expertise, which is assessing and handling customers' credit risk. The customer repays the loan and assumes the investment repayment risk (the financing institution [FI], of course, always has some risk for loan nonpayment. The assessment of a customer's credit risk is done by the FI; it is one of the factors that define interest rates). If the savings are not enough to cover debt service, the ESCO has to cover the difference. If savings exceed the guaranteed level, the customer typically pays an agreed-upon percentage of the savings to the ESCO (however, changes in energy consumption (e.g., business expansion or changes of processes or production lines) are likely to bring increased energy that can deteriorate the targets. Conversely, a contraction of business (e.g., an empty wing of a hotel) or a smaller production output will result in energy savings. Therefore, crucial issues to consider involve setting the baselines and associated growth projections, setting the system boundary and conditions, and preventing leakages. A clause in the contract that allows either party to reopen and renegotiate the baseline to reflect current conditions can solve this problem).

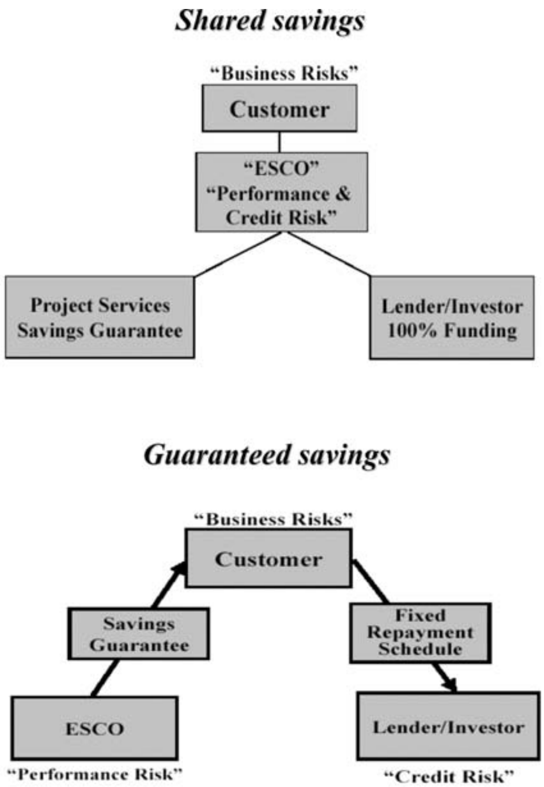


Fig. 2 Major types of performance contracting models/ repayment options. Under a shared savings contract, the cost savings are split for a predetermined length of time in accordance with a prearranged percentage. There is no standard split, as this depends on the cost of the project, the length of the contract, and the risks taken by the energy service company (ESCO) and the consumer. Under a guaranteed savings contract, the ESCO guarantees a certain level of energy savings and in this way shields the client from any performance risk.
Source: From Joint Research Centre, European Commission (see Ref. 4).

Energy Eff–Serv

Usually, the contract also contains a proviso that the guarantee is good (i.e., the value of the energy saved will be enough to meet the customer's debt obligation) provided that the price of energy does not go below a stipulated floor price (performance contracting is risk management, and dropping fuel prices—such as those experienced in North America in 1986—gave rise to this provision. We are indebted to Shirley Hansen for this clarification).[8] A variation on guaranteed savings contracts is pay-from-savings contracts, whereby the payment schedule is based on the level (percentage) of savings: the more the savings, the quicker the repayment.[13]

The guaranteed savings scheme is likely to function properly only in countries with an established banking structure, high degree of familiarity with project financing, and sufficient technical expertise, also within the banking sector, to understand energy-efficiency projects (e.g., in the United Kingdom; Austria; and, more recently, Hungary). The guaranteed savings concept is difficult to use in introducing the ESCO concept in developing markets because it requires customers to assume investment repayment risk. However, guaranteed savings foster long-term growth of ESCO and finance industries.[4] Newly-established ESCOs with no credit histories and limited resources would be unable to invest in the projects they recommend and might enter the market only if they guarantee the savings and the client secures the financing on its own.

In the United States, the guaranteed savings model evolved from the shared savings model in response to customers' desire to significantly reduce financing costs in exchange for accepting more risk due to their increased comfort with energy savings technologies. It also was initiated by smaller ESCOs and fostered by financial institutions to allow them to grow their respective industries. The primary benefit of this structure is that its reduced financing cost enables many more project investments to be made for the same debt service level. The public sector normally prefers this structure to maximize the amount of infrastructure investment made in its facilities from an EPC.

Conversely, under a shared savings model, the ESCO assumes both performance and credit risk. This is why a shared savings contract is more likely to be linked with TPF; ESCO financing; or a mixed scheme, with financing coming from the client and the ESCO whereby the ESCO repays the loan and takes over the credit risk. The ESCO therefore assumes both performance and the underlying customer credit risk; if the customer goes out of business, the revenue stream from the project will stop, putting the ESCO at risk.[13] Unfortunately, such a contractual arrangement may create leveraging and increased capital requirement problems for ESCOs because ESCOs become too indebted, and at some point financial institutions may refuse lending to an ESCO due to a high debt-to-equity ratio (experience in the United States shows that lenders tend to require a variety of credit enhancements for this type of financing, such as bonding or insurance[13]). In effect, the ESCO collateralizes the loan with anticipated savings payments from the customer based on a share of the energy cost savings. The financing in this case goes off the customer's balance sheet (under off-balance-sheet financing, also called nonappropriation financing, financiers hold title to equipment during the term of the agreement. Furthermore, to avoid the risk of energy price changes, it is possible to stipulate in the contract a single energy price. In this situation, the customer and the ESCO agree on the value of the service up front, and neither side gains from changes in energy prices. If the actual prices are lower than the stipulated floor value, the consumer has a windfall profit, which compensates the lower return of the project; conversely, if the actual prices are higher than the stipulated ceiling, the return on the project is higher than projected, but the consumer pays no more for the project. In effect, this variation sets performance in physical terms with fixed energy prices, which makes the approach resemble guaranteed savings[11]). A situation where savings exceed expectations should be taken into account in a shared savings contract. This setting may create an adversarial relationship between the ESCO and the customer[7] whereby the ESCO may attempt to lowball the savings estimate and then receive more from the excess savings (deliberate estimation of the lower value of savings is not restricted to the shared savings model; it is a standard practice for the ESCO to secure itself for the guaranteed performance with some buffer. The real questions are how big this buffer/cushion is and how the "excess" savings above the estimated ones are split between the client and the ESCO).[11]

The shared savings concept is a good introductory model in developing markets because customers assume no financial risk (the customers may have different reasons to be reluctant to assume financing, even if the cost of capital is higher for ESCOs than for customers. Among the reasons are adversity to assuming debt, borrowing limits, and budgetary restraints[11]). From the ESCO's perspective, the shared savings approach has the added value of the financing service.[11] However, this model tends to create barriers for small companies; small ESCOs that implement projects based on shared savings rapidly become too highly leveraged and unable to contract further debt for subsequent projects.[8,11] Shared savings concepts therefore may limit long-term market growth and competition between ESCOs and between FIs. For instance, small or new ESCOs with no previous experience in borrowing and with few resources are unlikely to enter the market if such agreements dominate.[3,4] It focuses the attention on projects with short payback times ("cream skimming"). Table 1 summarizes the features of the guaranteed and shared savings models.

Table 1 Guaranteed savings and shared savings: a comparison

Guaranteed savings	Shared savings
Performance related to level of energy saved	Performance related to cost of energy saved
Value of energy saved is guaranteed to meet debt service obligations down to a floor price	Value of payments to energy service companies (ESCO) is linked to energy price
Energy service companies carries performance risk energy user/customer carries credit risk	Energy service companies carries performance and credit risk as it typically carries out the financing
If the energy user/customer borrows, then debt appears on its balance sheet	Usually off the balance sheet of energy user/customer
Requires creditworthy customer	Can serve customers that do not have access to financing, but still requires a creditworthy customer
Extensive measurement and verification (M&V)	Extensive M&V
Energy service companies can do more projects without getting highly leveraged	Favors large ESCOs; small ESCOs become too leveraged to do more projects
More comprehensive project scope due to lower financing costs	Favors projects with short payback ('cream skimming') due to higher financing costs

Source: From Refs. 4,7,8 and 11.

Other Contracting Models

While there are numerous ways to structure a contract—hence, any attempt to be comprehensive in describing contracting variations is doomed—other contractual arrangements deserve some attention. Here, we describe the chauffage contract, the first-out contract, the build-own-operate-transfer (BOOT) contract, and the leasing contract.

A very frequently used type of contract in Europe is the chauffage contract, in which an ESPC or an ESCO takes over complete responsibility for the provision to the client of an agreed set of energy services (e.g., space heat, lighting, motive power). This arrangement is a type of supply-and-demand contract, and in effect it is an extreme form of energy management outsourcing. Where the energy supply market is competitive, the ESCO in a chauffage arrangement also takes over full responsibility for fuel/electricity purchasing. The fee paid by the client under a chauffage arrangement is calculated on the basis of its existing energy bill minus a percentage saving (often in the range of 5%–10%), or a fee may be charged per square meter of conditioned space. Thus, the client is guaranteed an immediate saving relative to its current bill. The ESCO takes on the responsibility of providing the improved level of energy service for a reduced bill. The more efficiently and cheaply it can do this, the greater its earnings. Chauffage contracts give the strongest incentive to ESPCs or ESCOs to provide services in an efficient way.

Chauffage contracts are typically very long (20–30 years), and the ESCO provides all the associated maintenance and operation during the contract. Chauffage contracts are very useful where the customer wants to outsource facility services and investment.[13] Such contracts may have an element of shared savings in addition to the guaranteed savings element to provide incentives for the customer. For instance, all savings up to an agreed figure would go to the ESCO to repay project costs and return on capital; above this, savings will be shared between the ESCO and the customer.

Another variation is the first-out approach, whereby the ESCO is paid 100% of the energy savings until the project costs—including the ESCO profit—are fully paid. The exact duration of the contract will actually depend on the level of savings achieved: the greater the savings, the shorter the contract.[5]

A BOOT model may involve an ESCO designing, building, financing, owning, and operating the equipment for a defined period of time and then transferring this ownership across to the client. This model resembles a special-purpose enterprise created for a particular project. Clients enter into long-term supply contracts with the BOOT operator and are charged accordingly for the service delivered. The service charge includes capital and operating-cost recovery and project profit. BOOT schemes are becoming an increasingly popular means of financing CHP projects in Europe.

Leasing can be an attractive alternative to borrowing because lease payments tend to be lower than loan payments. This method is commonly used for industrial equipment. The client (lessee) makes payments of principal and interest, and the frequency of the payments depends on the contract. The stream of income from the cost savings covers the lease payment. The ESCO can bid out and arrange an equipment lease-purchase agreement with a FI. If the ESCO is not affiliated with an equipment manufacturer or supplier, it can bid out, make suppliers' competitive analysis, and arrange the equipment.

There are two major types of leases: capital and operating. Capital leases are installment purchases of equipment. In a capital lease, the lessee owns and depreciates the equipment and may benefit from associated tax benefits. A capital asset and associated liability appears on the balance sheet. In an operating lease, the owner of the asset (lessor—the ESCO) owns the equipment and essentially rents it to the lessee for a fixed monthly fee. This is an off-balance-sheet financing source. It shifts the risk from the lessee to the lessor but tends to be more expensive for the lessor. Unlike in capital leases, the lessor claims any tax benefits associated with the depreciation of the equipment. The nonappropriation clause means that the financing is not seen as debt.

CONCLUSION

This entry has provided a concise overview of the key terminology related to energy service provision in Europe. It is the belief of the authors that the existence of standard, undisputed, and commonly used terms will facilitate the demand for, and ultimately the deployment of, more energy services.

An important step to securing the long-term credibility of energy efficiency is the ability to verify reduced consumption. To help end users and the financial community better understand EPC and gain confidence in the return on investment, it would be extremely beneficial to standardize savings M&V procedures. While the development of standard M&V has been an elusive task in Europe, as various companies consider their approaches unique and proprietary rather than developing a single standard, energy-service agreement efforts may be channeled toward agreeing upon a standard language for a set of key contract provisions, such as insurance, equipment ownership, and purchase options, which will allow standard contract forms to be built up gradually.

Due to space limitations, this entry has presented a fraction of the findings of the ongoing survey of ESCOs in Europe. Further information—including a concise review of the current status of the ESCO industry in Europe—on the project specifics of the most common types of activities and of the features of the ESCO industry in selected EU member states is available, together with an outline of the key factors of success of national ESCO industries and a list of suggested strategic actions needed to increase the deployment of ESCOs in Europe.[1]

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions toward working out often dubiously interpreted ESCO terminology for the "ESCOs in Europe" status report 2005 of Shirley Hansen, Thomas Dreessen, Pierre Langlois, and Anees Iqbal. The authors would like to thank all national and international experts who provided input to the survey and responded to our questions—and, of course, all ESCOs that provided information that was used for the analysis in the "ESCOs in Europe" report.

REFERENCES

1. Bertoldi, P.; Rezessy, S. *Energy Service Companies in Europe: Status Report 2005*. Published by DG Joint Research Centre of the European Commission and available for download at http://energyefficiency.jrc.cec.eu.int/pdf/ESCO%20report%20final%20revised%20v2.pdf, (accessed on December 2006).
2. Bertoldi, P.; Berrutto, V.; De Renzio, M.; Adnot, J.; Vine, E. *How are ESCOs behaving and how to create a real ESCO market*. In Proceedings of the European Council for Energy Efficient Economy 2003 Summer Study; European Council for an Energy-Efficient Economy: Stockholm, 2003.
3. Butson, J. In *The Potential for Energy Service Companies in the European Union*, Proceedings of the First International Conference on Improving Electricity Efficiency in Commercial Buildings, Amsterdam, September 1998.
4. CTI (Climate Technology Initiative). *Guide to Working with Energy Service Companies in Central Europe*; CTI Secretariat: Tokyo, 2003.
5. Dreessen, T. In *Advantages and Disadvantages of the Two Dominant World ESCO Models; Shared Savings and Guaranteed Savings*, Proceedings of the First Pan-European Conference on Energy Service Companies; Bertoldi, P., Ed.; Joint Research Centre, European Commission, 2003.
6. ECS (Energy Charter Secretariat). Third party financing. Achieving its Potential; Energy Charter Secretariat: Brussels, 2003.
7. Goldman, C.; Hopper, N.; Osborn, J. Review of U.S. ESCO industry market trends: an empirical analysis of project data. Energy Policy, **2005**, *33*, 387–405.
8. Hansen, S. *In Performance Contracting Models and Risk Management*, Presentation at the workshop developing the business of energy efficiency performance contracting organized by Ibmec and INEE, Rio de Janeiro, 2003.
9. Hansen, S. Personal Communication; Kiona International: U.S.A., 2004.
10. Hansen, S.; Brown, J. *Investment Grade Audit: Making Smart Energy Choices*; Fairmont Press: Lilbourn, 2003.
11. McGowan, J. State-of-the-art performance contracting. Article available online at http://www.energyusernews.com/CDA/ArticleInformation/features/BNP__Features__Item/0,2584,61014,00.html (accessed on December 2005).
12. Poole, A.D.; Stoner, T.H. Alternative financing models for energy efficiency performance contracting. *Sponsored by the USAID Brazilian Clean and Efficient Energy Program (BCEEP)*; 2003; administered by Winrock International.
13. Sorrell, S. The economics of energy service contracts. *Tyndall Centre Working Paper*, Environment and Energy Programme SPRU (Science and Technology Policy Research): Freeman Centre, 2005.
14. WEEA, World Energy Efficiency Association. *Briefing Paper on Energy Service Companies with Directory of Active Companies*; WEEA: Washington, 1999.

Encyclopedia of

Energy Engineering and Technology

First Edition

Volume 2

Pages 573 through 1146
Energy Star–Perf

Energy Star® Portfolio Manager and Building Labeling Program

Bill Allemon
North American Energy Efficiency, Ford land, Dearborn, Michigan, U.S.A.

Abstract
This entry on the ENERGY STAR Portfolio Manager and building-labeling program acts as a single resource for first-time users of this energy benchmarking application from the U.S. Environmental Protection Agency's ENERGY STAR program. The entry provides background on the need for energy benchmarking, development of the Portfolio Manager application, and typical uses. Details regarding applicability of the tool, benchmarking requirements, and energy data normalization are included. The entry reviews the ENERGY STAR building-labeling program and requirements to apply for the building label. Last, additional uses of the application's results and additional related ENERGY STAR tools are included.

INTRODUCTION

Demand for accurate energy efficiency benchmarking increases as energy, labor, and other facility costs continue to rise. A number of benchmarking tools exist to access the energy efficiency of select facilities, but few are as comprehensive and widely adopted as Portfolio Manager, available from the U.S. Environmental Protection Agency (EPA)'s ENERGY STAR program.

This entry provides a general overview of the uses and applicability of Portfolio Manager. It is not intended to be a comprehensive user's guide explaining each individual data input screen or to be a technical narrative of the tool's statistical analysis method; neither does it attempt to act as a replacement for the extensive support and resources available through ENERGY STAR. It provides the user a solid understanding of the uses and benefits of the tool, offers tips on how to get started, and acts as a guide through the vast library of support materials and documents available at the ENERGY STAR web site. If there is one area of weakness in Portfolio Manager, it is the lack of a single, comprehensive user's manual.

BACKGROUND

ENERGY STAR is a program funded by the U.S. government and administered by the EPA, in partnership with the U.S. Department of Energy (DOE), to reduce the nation's consumption of energy by facilities and industrial processes. The program's main goal is to lower significantly the emission of greenhouse gases—specifically carbon dioxide (CO_2)—by reducing energy consumption. ENERGY STAR provides tools, resources, and support to a wide variety of markets, including residential, commercial, retail, industrial, and many others. ENERGY STAR is the leading contributor to the development and establishment of energy efficiency standards for home appliances; most U.S. consumers recognize the ENERGY STAR logo from retail shopping. The ENERGY STAR program facilitates a number of industry focus groups, in which manufacturers can come together in a noncompetitive environment to improve the energy efficiency of their industrial sector. The success of these focus groups is evidenced by the number of industry-specific tools and best practices generated by these sessions. More information on industry-specific groups can be found on the ENERGY STAR Web site.

The Need for Energy Benchmarking

Rising energy costs and increased competition are driving facility and property managers to evaluate every aspect of fixed and variable costs. Having a common measure of efficiency is essential so that all users within an industry are using the same standard of measure. Benchmarking tools must be accurate, recognized, and accepted by each industry, and must use a common industry baseline. The means and methods used in normalizing and analyzing data must also be standardized; otherwise, each user will have an inaccurate representation of his or her facility's efficiency relative to the industry as a whole.

Obtaining accurate competitive data can be extremely difficult, as is finding applicable energy modeling methods. Even if the user can gain access to competitive data, several methods exist to analyze facility data and normalize variances. A variety of methods exist to filter out the effect of significant influencers, such as weather, high-intensity

Keywords: ENERGY STAR; Building; Label; Certification; Benchmarking; EPA; Portfolio manager; Commercial.

loads, and facility use. When comparing one facility with others, it's important to use consistent analysis methodology to ensure predictable and accurate results. Each statistical model has its own limitations and caveats. An excellent technical reference regarding energy modeling is Measuring Plant-Wide Energy Savings, by Dr Kelly Kissock and Carl Edgar of the University of Dayton.[1]

After the issue of benchmarking accuracy is addressed, its applicability and many uses can be discussed. Benchmarking can identify whether energy is being wasted and, if so, how much. Quantifying the amount of energy waste provides a valuable number to the facility manager; it tells the manager up front how much load must be removed for the building to reach an identified level of efficiency. Without this value, a facility manager must find the value through trial and error, reducing consumption wherever he or she can, while waiting until a new set of energy data is collected and the results are analyzed.

Buildings can also be benchmarked against themselves to identify whether efficiency is improving, staying within limits, or getting worse. This method can also identify whether an energy reduction project has indeed met its goal and then quantify the actual savings. This feature is helpful for validating the effectiveness of energy performance contract projects.

Last, most companies and organizations do not have the level of in-house expertise required to model energy consumption accurately. Portfolio Manager is easy enough for a nontechnical manager to use without needing the engineering and statistical knowledge required to build a predictive model from scratch.

In summary, facility managers need a standardized and industry-accepted tool to enable the accurate identification of facility energy efficiency compared with that of similar buildings across the United States. The ENERGY STAR Portfolio Manager application addresses this need.

PORTFOLIO MANAGER OVERVIEW

Portfolio Manager is an online software application that benchmarks the energy use of a facility against a national database of similar buildings. In use since 1999, Portfolio Manager is an industry-accepted energy use model that was developed exclusively for ENERGY STAR by the University of Dayton. More than 26,000 buildings and more than 4 trillion square feet have been evaluated using Portfolio Manager, with more than 2500 buildings obtaining the ENERGY STAR label. The Portfolio Manager database continues to grow in size and accuracy as registered partners enter consumption data and analyze facilities, and usage continues to increase. Within the past 3 years, the number of architectural engineering firms using Portfolio Manager has increased by a factor of four.[2]

A variety of building types is included in the national database, representing more than 50% of the office space located within the United States.[3] Those building types include

- General office buildings
- Financial centers
- Bank branches
- Courthouses
- K-12 schools
- Hospitals
- Hotels and motels
- Supermarkets
- Residence halls/dormitories
- Warehouses

As Portfolio Manager analyzes the energy use of a subject facility, it adjusts or normalizes certain factors to compare facilities accurately. Adjustments can compensate for building size, geographic location, year-to-year weather variances, and other parameters.

Output from the model is expressed as a 0–100 percentile ranking, in which 50% identifies average energy efficiency (as compared with the national database) and 75% or higher identifies an energy-efficient facility. Data and results from a particular facility or group of buildings can be shared in a controlled fashion with other users. A number of energy efficiency statistics are generated by the tool, including overall energy intensity, the facility's Energy Performance Rating (0–100 percentile ranking), and effective CO_2 emissions. These results can be generated for one facility or a group of buildings.

Output from the model can be used to determine whether additional investment in the facility's infrastructure and systems is worthwhile or whether investment dollars should be spent at another location. Building ranking can also be used to quantify the improvement in efficiency and cost after energy-saving projects have been installed. Lastly, output from Portfolio Manager may be used to apply for the ENERGY STAR building label, if the facility has achieved a score of 75 or higher. Buildings that meet or exceed this level of efficiency, and that meet other requirements for indoor comfort and safety, will receive an ENERGY STAR plaque to display within or to affix to the benchmarked facility. This indicates to all that the facility resides within the top 25% of energy-efficient facilities within the United States.

USING PORTFOLIO MANAGER

Getting Started

ENERGY STAR provides new users a variety of ways to become familiar with Portfolio Manager. These include a self-guided online tour, regularly scheduled Web-based conferences, and a large collection of support documentation. These resources help the user understand how to apply the tool and the resultant output, and what the input data requirements are for specific facility types. This section will guide the user through the process of determining whether a specific facility can be benchmarked using the tool, identify necessary input data, review weather normalization methodology, and explain additional requirements of specific facility types.

Eligibility Requirements

For Portfolio Manager results to be valid, specific requirements have been established to ensure that the facility being benchmarked is similar to those contained within the tool's nationwide database. These requirements are called Eligibility Rulesets. Definitions exist for each facility and space type.

These requirements call for at least 50% of a facility's square-footage floor space to match the rulesets for each space type and for at least 11 months of energy use data to be entered. These requirements and links to others can be found on the online page Eligibility Requirements for Use of the National Energy Performance Rating System, located on the ENERGY STAR web site.[4]

Following is an overview of the general requirements of each ruleset. Before applying for the ENERGY STAR building label, the user should review the current rulesets and definitions at the ENERGY STAR web site.

- *General office buildings*. The building is used for administrative, professional, or general office purposes. Supporting functions—conference rooms, lobbies, stairways, etc.—should be included in the total. The building must operate at least 35 h/week and cannot contain more than 25,000 occupants. The total building size must be at least 5000 ft^2.
- *Financial centers*. The facility is used as a bank headquarters or as a securities or brokerage firm. The total building size must be at least 20,000 ft^2.
- *Bank branches*. The facility is used for bank-branch-office purposes. The total building size must be at least 1000 ft^2.
- *Courthouses*. The facility is used for federal, state, or local courts and associated office space. The total building size must be at least 5000 ft^2.
- *K-12 schools*. School buildings for kindergarten through 12th grade. The total building size must be at least 5000 ft^2.
- *Hospitals*. The facility must have at least 50% of its total square-footage floor space dedicated to acute care. These are facilities that typically provide a variety of services within the same building or among multiple buildings on a campus, including emergency medical care, physician's-office care, diagnostic care, ambulatory care, and surgical care. The facility must contain at least 16 licensed beds and may not exceed 1510 licensed beds. The building cannot exceed 40 floors. The total building size must be at least 20,000 ft^2 but not more than 5 million ft^2. Children's hospitals can also be benchmarked using the tool.
- *Hotels and motels*. Facilities that rent overnight accommodations with a bath/shower and other facilities in most guest rooms. Floor area for all supporting functions such as food facilities, laundry facilities, and exercise rooms should be in the total. Hotel/motel categories currently eligible for benchmarking include economy, midscale, upscale, and upper upscale. Resort and extended-stay categories are not eligible for benchmarking at this time. Total building size must be at least 5000 ft^2.
- *Supermarkets*. Facilities used for the retail sale of food and beverage products. Entire floor space—including warehouse, office, break room, and storage areas—should be included. Total building size must be at least 5000 ft^2.
- *Residence halls/dormitories*. Buildings that offer multiple accommodations for long-term residents associated with educational institutions or military facilities. Floor area for all supporting functions—food facilities, laundry facilities, exercise rooms, health club/spas, and others—should be included. Total building size must be at least 5000 ft^2.
- *Warehouses*. Facility space used to store goods, manufactured products, merchandise, or raw materials in a refrigerated or nonrefrigerated building. Refrigerated warehouses are those facilities designed to store perishable goods or merchandise at temperatures below 50°F. Nonrefrigerated warehouse is space designed to store nonperishable goods and merchandise. The total building size must be at least 5000 ft^2.

In addition to the primary spaces noted above, secondary space uses can be analyzed by Portfolio Manager. For each of these secondary space types, weekly hours of operation and gross square footage must be defined. Typically, the aggregate of the secondary space types cannot represent more than 50% of the total square footage of a building, and a facility cannot be comprised solely of secondary space types.

A few examples of secondary space types and their specific requirements are given below. Additional information regarding secondary space types can be found at the online page Eligibility Requirements for Use of the National Energy Performance Rating System, located on the ENERGY STAR web site.[4]

- *Ambulatory surgical centers*. These facilities are categorized by size and weekly hours. This space cannot be more than 50% of the facility's total floor area. The gross floor square footage cannot exceed 1 million square feet.
- *Computer data centers*. These facilities are designed and equipped to meet the needs of computer equipment

that must reside in areas with independently controlled temperature and humidity. Usually, the air conditioning system for this type of area is separate from the equipment used to control the space temperature of the general building. The total area of all computer data center spaces cannot exceed 10% of the total building space.

- *Garages and parking lots*. Enclosed or open parking facilities that operate from the same energy use meter as the primary building. Garages are categorized as either above ground or below ground. Below-ground garages are assumed to have mechanical ventilation. All parking facilities are assumed to have some kind of lighting. Garage space cannot account for more than 50% of the total building area.

As with any requirements and definitions, the user should first obtain the complete and most recent rulesets by visiting the ENERGY STAR web site.

Information Required for Data Input

Two levels of input data are required for facilities benchmarked by Portfolio Manager: facility summary data and facility space information. When the user creates a new building file, the application will ask for basic information, such as the building address, space use, energy use, and baseline/target usage.

Depending on the type of building being evaluated, Portfolio Manager will require specific input parameters to be identified. These include floor area, the number of personal computers, occupancy levels, and operating hours. Specific facility types will require additional information to benchmark the facility accurately. K-12 schools, for example, are required to supply the following additional information: the percentage of floor space that is heated, the percentage of floor space that is air conditioned, the square footage of the cooking area, the months that the school is in operation, the number of students, and whether the building has mechanical ventilation. Each of these parameters has been identified as a significant influence on the energy use profile of K-12 schools. Other facility types require specific input data relative to their particular drivers of energy consumption.

WEATHER NORMALIZATION

As can be expected, the geographic location of a facility greatly affects its energy consumption. A refrigerated warehouse designed to maintain an internal temperature of 45°F, for example, will have significantly different energy usage in Minnesota, Nevada, or California. When the need to control and maintain humidity levels is added, energy usage varies even more greatly. One additional factor includes severe weather variations at a specific location.

The statistical model used within Portfolio Manager normalizes changes in both historical temperature variances and year over year severe or mild temperature changes. Publicly available weather data are used to make these adjustments. Heating-degree-day (HDD) and cooling-degree-day (CDD) values are indices used to represent the energy demand placed on building heating and cooling equipment. The values are calculated from individual daily temperatures for a specific region or U.S. postal code.[5]

The statistical model used by Portfolio Manager is based on the E-Tracker software model developed by Dr Kelly Kissock of the University of Dayton. Additional technical information and an outline of the weather normalization methodology used by this statistical model can be found in the document Weather Normalization Description, located on the ENERGY STAR web site.[5]

THE ENERGY STAR BUILDING LABEL

Overview

Buildings with a Portfolio Manager score of 75 or higher that meet specific indoor air quality standards verified by a licensed professional engineer are eligible to apply for the ENERGY STAR building label. The label signifies the relative energy efficiency of the facility compared with Portfolio Manager's national building database. The user is required to generate a Statement of Energy Performance (SEP) using Portfolio Manager indicating a score of 75 or higher, have the input data and other requirements validated by a licensed professional engineer (PE), and submit a letter of agreement and the SEP (complete with PE stamp) to the EPA for review.

Statement of Energy Performance

The SEP validates the energy efficiency of a building. Facility managers can leverage the information in this document in a number of ways. Proof of superior energy efficiency can be beneficial during the sale or lease of a facility, confirming that facility improvements have been made. The document can also be used to demonstrate the value of an organization's energy efficiency program.

In addition to the energy performance rating of the facility, information in the SEP includes the year the facility was built, gross square footage, ownership and contact information, energy use summaries, and intensities with CO_2-emissions equivalents. The SEP has an area to hold a PE stamp authenticating the information in the document.

Professional Engineer Review and Validation

ENERGY STAR requires that a licensed PE verify that the data entered into Portfolio Manager identifying the building energy consumption and physical parameters are accurate. The engineer must visit the building and verify that it conforms to current industry standards for indoor environment, as referenced by ENERGY STAR in the Professional Engineer's Guide.[6] These standards address indoor air quality, illumination, and occupant comfort, as identified by the American Society of Heating, Refrigerating, and Air-Conditioning Engineer's (ASHRAE).

ENERGY STAR relies on the professionalism of the PE industry to provide an unbiased, trained, and accurate review of the building performance data. Because PEs are legally bound to uphold their standard of ethics, the PE stamp brings a mark of credibility and accuracy to the SEP. The PE stamp makes the SEP valid and official, whether the SEP is being used to apply for a building label or for any other purpose.

The Professional Engineer's Guide to the ENERGY STAR Label for Buildings is the technical reference used by PEs when validating data.[6] The guide has a chapter dedicated to each category or performance characteristic. Each chapter is clearly written in an outline format, identifying its purpose, identifying any reference standards, and defining acceptable conditions. Each chapter also outlines expectations of the PE, provides hints and tips, and concludes with a section on commonly asked questions and related answers. Overall, the guide is a statement of work identifying the responsibilities and expectations of the PE.

Indoor Air Quality Requirements

One common method used to improve energy efficiency, especially in commercial buildings, is reducing the quantity of airflow generated by heating and ventilating (H&V) equipment. Many commercial buildings are indeed overventilated due to poor commissioning, lack of maintenance, or conservative system design. However, lowering air flow reduces the volume of fresh outdoor air brought into a facility, thereby reducing indoor air exchanges—the number of times a facility's indoor air is replaced. Failure to properly exhaust used indoor air and its contaminants (including CO_2, mold, mildew, and microbes) can result in a number of health complaints and issues. To ensure that energy efficiency isn't being achieved at the expense of indoor air quality (IAQ), ENERGY STAR has identified a list of requirements to balance the goal of improving efficiency while maintaining a safe and healthy work environment.

An IAQ assessment is to be performed by the PE during the on-site building evaluation. The engineer is required to conduct a point-in-time assessment of the building's IAQ against four industry accepted standards, including:

- ASHRAE Standard 62-1999—Ventilation for Acceptable Indoor Air Quality
- ASHRAE Standard 55-2004—Thermal Environmental Conditions for Human Occupancy
- ASHRAE Standard 52.1-1992—Gravimetric and Dust Spot Procedures for Testing Air Cleaning Devices Used in General Ventilation for Removing Particulate Matter
- 9th Edition of the Illuminating Engineering Society of North America Lighting Handbook

Applying for the Building Label

Prior to applying for the ENERGY STAR building label, the user should review the Eligibility Rulesets for the building being benchmarked and use Portfolio Manager to identify an Energy Performance Rating for the facility.[7] If the rating is 75 or higher, the following additional requirements must be met.

Using the Professional Engineer's Guide, a licensed PE must verify that the building meets industry standards for occupant comfort and IAQ.[6] The PE must be licensed in the state where the building resides and will be required to stamp and sign the SEP. The facility manager should also read and understand the ENERGY STAR Identity Guidelines.[8] This booklet identifies usage requirements for the ENERGY STAR logo and brand name.

Finally, a signed letter of agreement is required. This document is also generated through Portfolio Manager. Mail the signed letter of agreement and the signed/stamped SEP to ENERGY STAR within 120 days of the energy consumption data's period ending date. Visit the ENERGY STAR Web site to verify the correct mailing address for the ENERGY STAR Label for Buildings program.

After final review of submitted forms and data, the ENERGY STAR program will award the facility a building plaque to display in a prominent area of the building.

ADDITIONAL PORTFOLIO MANAGER TOOLS

Target Finder

Target Finder is a Web-based software tool that generates a predictive Portfolio Manager score for a facility during its design stage. Using this tool, engineers and architects can establish an energy use target for the facility, enter estimated annual energy use, and then compare the building profile against the same database used by Portfolio Manager. Similar to Portfolio Manager, the tool will adjust estimated energy consumption based on space type and geographic location. The tool will not only generate an energy performance rating based on inputs, but also generate predictive energy use values for a target performance rating. Using this feature, the designer can not only forecast a performance rating, but perform a gap analysis to identify the quantity and type of fuel to be reduced in order to achieve the target score.

Delta Score Estimator

Similar to Target Finder but much less extensive, this tool quickly and easily quantifies the reduction in energy consumption necessary to improve a facility's Portfolio Manager score above the 75% level. The tool can also be used to identify a projected score, given a proposed reduction in consumption. Use this tool when analyzing existing facilities that do not generate an energy performance rating high enough to apply for a ENERGY STAR building label.

Tracking Water Consumption

In 2006, Portfolio Manager was enhanced to permit tracking of facility water consumption. Because commercial buildings consume almost 20% of U.S. drinking-water supplies, a reduction of only 10% would equate to the conservation of more than 2 trillion gal of water each year. Portfolio Manager now allows the facility to track water consumption in four major categories: Indoor Use, Outdoor Use, Combined Indoor/Outdoor Use, and Wastewater.[9] The new feature permits benchmarking water consumption over time and enables relationships between energy use and water consumption to be identified.

ENERGY STAR plans to expand this feature further by enabling Portfolio Manager to quantify water consumption per occupant and develop a percentile ranking system for water consumption.

WHERE TO GO FROM HERE

A key component of the ENERGY STAR program is the act of sharing best practices and lessons learned with others. This interaction is the primary element that sustains and strengthens the program over time, while the number of users and registered partners continues to grow. ENERGY STAR has a number of programs that users can employ to share experiences and expand their own knowledge base. A few are highlighted below.

Building Profiles

ENERGY STAR Building Profiles are case studies outlining the steps taken by an organization to improve the energy efficiency of a subject facility during its path toward obtaining a building label. Building Profiles are created within Portfolio Manager and submitted to ENERGY STAR for review from within the tool. After final review and approval by an EPA committee, the Building Profile is made available to all via the ENERGY STAR Registry of Labeled Buildings. First-time users may find it helpful to review the existing Building Profiles of similar facilities before attempting to achieve a Portfolio Manager score and embarking on making costly facility upgrades.

Industry-Specific Energy Performance Indicators

ENERGY STAR champions a number of industry-specific focus groups to give large companies the opportunity to improve the environment while sharing best practices. One of the outcomes of these sessions is the Energy Performance Indicator (EPI). Initially generated for the automotive industry, this successful tool is being replicated for other industries, including cement manufacturing, wet-corn milling, glass manufacturing, and food processing plants. Similar to the Portfolio Manager tool, the EPI benchmarks energy consumption for a specific industrial facility against similar facilities across the United States while normalizing energy consumption against statistically significant influencers. The automotive EPI, for example, normalizes motor-vehicle assembly plant consumption against variations in weather, production volume, conveyor line speed, vehicle wheelbase, and plant air tempering. For more information on industrial EPI tools, visit the Industries in Focus page on the ENERGY STAR web site.[10]

ACKNOWLEDGMENTS

The author would like to thank the ENERGY STAR staff at the EPA, the DOE, and the ENERGY STAR team of consultants and contractors for their assistance in teaching, applying, and supporting Portfolio Manager and its related tools. Without the extensive support system and network of registered ENERGY STAR partners, the tools outlined in this entry would significantly lose their impact and usefulness.

REFERENCES

1. Eger, C.; Kissock, K. Measuring plant-wide energy savings. In *SAE Technical Paper Series*, 2006 SAE World Congress, Detroit, Michigan, April 3–6, 2006; SAE International: Detroit, Michigan, 2006; 2006-01-0596.
2. Available at www.energystar.gov/ia/news/downloads/2005_achievements.pdf (accessed September 2006).
3. Available at www.energystar.gov/index.cfm?c=evaluate_performance.bus_portfoliomanager (accessed September 2006).
4. Available at www.energystar.gov/index.cfm?c=eligibility.bus_portfoliomanager_eligibility (accessed September 2006).
5. Available at www.energystar.gov/index.cfm?c=business.bus_weather_normalization (accessed September 2006).

6. U.S. Environmental Protection Agency, Office of Air and Radiation. *Professional Engineer's Guide to the ENERGY STAR Label for Buildings*, EPA 430-F-01-XX; EPA: Washington, DC, 2003.
7. Available at www.energystar.gov/index.cfm?c=evaluate_performance.bus_portfoliomanager_intro (accessed September 2006).
8. Available at https://www.energystar.gov/index.cfm?c=logos.pt_guidelines (accessed September 2006).
9. Available at www.energystar.gov/index.cfm?c=business.bus_water (accessed September 2006).
10. Available at www.energystar.gov/index.cfm?c=in_focus.bus_industries_focus (accessed September 2006).

Energy Use: U.S. Overview

Dwight K. French
Energy Consumption Division, Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.

Abstract
Virtually all of the myriad economic and social activities of United States society involve inputs of energy. Our high activity levels have propelled the United States to the forefront of energy consumers among nations. Our almost 100 quadrillion (100 followed by 15 zeroes) British thermal units (Btu) of energy consumed annually is more than twice as much as that of any other nation. However, relative to our population and economic activity, the U.S. is comparable to, and even below, other nations in energy use. A key to understanding the demand for energy in the United States is comprehending the vast scope of activity to which we intensively devote ourselves. This article attempts to capture that scope by providing an overview of energy use in the several major sectors of U.S. society—housing, commerce and institutions, industry, transport, and even energy transformation—in terms of the types of energy used and the purposes for which it is used.

INTRODUCTION

Any characterization of United States society and its people would have to include the descriptor "active." The great majority of people, businesses, and institutions of this country are about "doing," filling our time with an enormous variety of activities and responsibilities. People's personal, social, and professional reputations are very much built on what we "do." Even much of our "leisure" time is devoted to the quest to keep us entertained, informed, and otherwise occupied.

Virtually all of this activity involves inputs of energy—not just human energy, but outside energy sources as well. Over the past 150 years, outside energy inputs have mushroomed in importance, becoming an indispensable part of our social and economic structure. Our high activity levels have propelled the United States to the forefront of energy consumers among nations. Our almost 1002 quadrillion (100 followed by 15 zeroes) British thermal units (Btu)[1] of energy consumed annually is more than twice as much as that of any other nation, though relative to our population and economic activity, the United States is comparable to, and even below, some other nations in energy use.[2] Only three other nations—China, Russia, and Japan—use more than 20 "quads" of energy in total.

A key to understanding the demand for energy in the United States is comprehending the vast scope of activity to which we intensely devote ourselves. This article attempts to capture that scope by providing an overview of energy use in the several major sectors of U.S. society—housing, commerce and institutions, industry, transport, and electricity production and delivery—as well as the types of energy used. The article then presents a more detailed description of each sector in terms of major energy forms used, the uses of that energy, and the outputs and outcomes that result. Most of the information in this article comes from the Energy Information Administration (EIA), the informational organization within the U.S. Department of Energy.

U.S. ENERGY USE ACROSS THE MAJOR SECTORS

One common initial way to partition U.S. energy use is to allocate it to five major activity groups (hereafter referred to as sectors): electricity generation, residential, commercial, industrial, and transportation. Electricity generation is often thought of as a transformation process that results in energy available for the other sectors rather than an end use of energy (that is, a use in which the outputs are nonenergy products or services). Residential use is, of course, energy use associated with housing units. Commercial use, as defined here, includes not only traditional economic commerce, but also use by other organizations and institutions, including civilian government—broadly, the service sector. Industrial use encompasses goods/products output: Manufacturing, agriculture, forestry, fisheries, mining, and construction. Transportation use encompasses all movement of people and goods, and thus is associated with the activity of all other sectors.

United States energy use can be partitioned by sector in three different ways, each with its own interpretation. First, the energy value of the electricity, along with all energy used up in its generation, transmission, and

Keywords: Commercial; Electricity generation; End use sectors; Energy consumption; Industrial; Residential; Transportation.

Table 1 Three approaches to distributing energy use by sector: United States, 2003 (quadrillion Btu)

Sector	All energy associated with generation	Usable electricity associated with end use	All energy associated with end use
All sectors	98.2	98.2	98.2
Generation	38.2	26.3	0.0
Residential	7.2	11.6	21.2
Commercial	4.2	8.4	17.5
Industrial	21.7	25.1	32.5
Transportation	26.8	26.8	26.9

Note: Components may not sum to totals because of independent rounding.
Source: From Energy Information Administration, Annual Energy Review, 2003.

distribution, can be allocated to the generation sector. This approach essentially treats electricity delivered to the other (so called "end use") sectors as a non-energy input. Second, the heat and power value of the electricity delivered to the end use sectors, rated at a standard 3412 Btu per kilowatt-hour (kWh),[3] can be allocated to those sectors, leaving the energy dissipated in the generation and distribution processes allocated to the generation sector. This approach treats the end use sectors as accountable for the use of outside-generated electricity as it reaches them and keeps the energy used up in generating and delivering electricity associated with the generation sector. Third, the total energy requirements for generating and delivering electricity can be allocated to the end use sectors, with no energy allocated to the generation sector. This approach treats the end use sectors, the ultimate energy consuming entities, as responsible for all of the energy embodied in the electricity they use, and not just the useful energy as it arrives at their doorsteps.

Table 1 shows that these three approaches paint much different pictures of relative energy requirements among the various sectors. Taking approach 1, the total energy requirements embodied in electricity generation in the United States far exceeds the energy requirements of any other sector; transportation is the largest consumer of the end use sectors; and the residential and commercial sectors combined account for only about one-ninth of energy use. Using approach 2, the generation, industrial, and transportation sectors account for almost equal shares (note that transportation uses very small amounts of electricity), and the residential and commercial sectors, the largest electricity users, represent 20% of energy use. Under approach 3, the generation sector disappears and the energy distribution among the four end use sectors is much more balanced, with industrial now accounting for the largest share of energy use.

This final approach clearly shows an underlying reason for U.S. predominance in energy use. We use large amounts of energy in all of our sectors. With electricity requirements allocated to the four major end use sectors, each of them alone would rank the United States no lower than fourth in energy use relative to the total energy use of other nations.

U.S. ENERGY USE BY TYPE OF ENERGY

Energy use in the United States is dominated by fossil fuel consumption. Fig. 1 shows that the three major fossil fuels, coal, natural gas, and petroleum, account for about 86% of the country's energy use. In terms of physical quantities, petroleum use is about 20 million 42-gallon barrels per day (enough to cover a regulation football field over 2000 ft deep), natural gas use is about 60 billion cubic feet per day (equivalent to a cubic mile of gas about every two and one-half days), and coal use, now almost exclusively for electricity generation, is about 2.7 million tons per day (enough to fill a train of standard 100-ton coal cars[4] over 300 miles long). Nuclear energy, used solely for electricity generation, accounts for another 8%, and a variety of renewable sources, primarily hydropower and biomass, comprise the remainder.

Solar, wind, and geothermal energy, renewable energy forms commonly considered to be the most environmentally appealing, together account for only about one-half of 1% of U.S. energy use at this time. The rate of increase in their use is much larger than the more conventional energy types, but they are starting from such a small base that many more years of rapid growth are required in order for them to become even a moderately important contributor to the U.S. energy mix.

ENERGY USE WITHIN SECTORS

Electricity Generation

Electricity generation is a huge enterprise in the United States. The amount of energy used just to generate our electricity ranks the United States a close second to the total energy use of China, and far ahead of the total energy

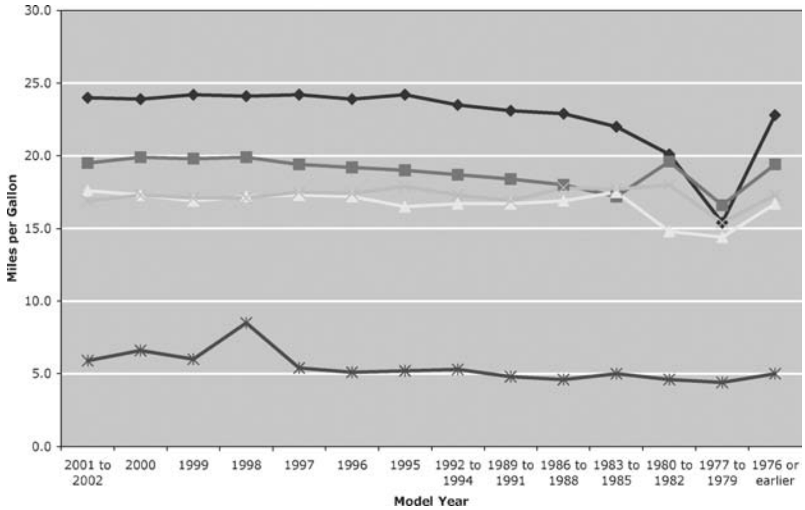


Fig. 1 Household vehicle fuel economy by vehicle type and model year. U.S., 2001.
Source: Reprinted with permission from Energy Information Administration (see Ref. 26).

requirements of any other country.[1,3,5] For many decades, virtually all electricity produced in the United States was generated within the traditional, vertically integrated electric utility, responsible for the entire process from initial generation to delivery to the final consumers. Over the past 30 years, various laws and regulations[6] have been put into place that encourage decentralized electricity production, and in some cases force the utilities to divest themselves of the generation process in order to deregulate the industry. As a result, the electric power industry looks much different than it did 30 years ago, or even a decade ago, in some ways.

Table 2 shows the distribution of electricity generated in the United States by type of generation facility and type of energy used in generation for the years 1993 and 2003. In 1993, electric utility generation accounted for the vast majority, 90.2%, of total generation, independent power producers accounted for 1.7%, and other generation, primarily commercial and industrial cogeneration, provided the remaining 8.1%. By 2003, the balance had notably shifted, with independent power producers claiming 27.4% share, carved out of the utility share. Much of this shift was due to divestiture of generating facilities by utilities as required by electricity deregulation, rather than closing of old facilities and construction of new ones. Thus, the distribution of generation by type of input energy would be expected to change much less dramatically than the distribution by type of facility, since ownership change by itself does not affect a facility's energy inputs.

And in fact, this was the case. Natural gas, the fuel of choice for a large part of new generation in the past decade, has made some limited inroads into generation share. However, the fundamental reliance on coal for electricity generation remains in place, use of nuclear energy has also grown, and hydro is in the same place as it was a decade earlier. This stability reflects a fundamental truth about U.S. energy use, both in this sector and, to a greater or lesser degree, across sectors: The country's energy demand infrastructure, built up over a century or more, will not change dramatically in the short term. Any significant technology-driven change in the country's energy consumption patterns will occur as the result of incremental changes that accumulate over a long time.

Table 2 Electricity generation by class of generator and input energy: United States, 1993 and 2003 (million mWh)

Generator class and input energy	1993	2003
Total generation	3197	3883
Class of generator		
Electric utility	2883	2462
Independent power producer	53	1063
Electric utility CHP[a]	108	196
Commercial/industrial CHP[a]	153	162
Input energy		
Coal	1690	1974
Petroleum	113	119
Natural gas	415	650
Nuclear energy	610	763
Hydro (net of pumped storage)	276	267
All other energy	93	109

Note: Components may not sum to total because of independent rounding.
[a]CHP: Combined Heat and Power.
Source: From Energy Information Administration, *Electric Power Annual,* Table 1.1 at http://www.eia.doe.gov/cneaf/electricity/epa/epat1p1.html (accessed May 2005).

Residential

The stock of housing units and their occupants currently account for about one-fifth of U.S. energy demand, including electricity requirements (Table 1). The Energy Information Administration's (EIA) Residential Energy Consumption Survey (RECS), which assesses energy use in the nation's occupied, year-round housing units, indicates that total residential energy use has remained remarkably stable over the past two-plus decades, even decreasing a bit, despite a 40% increase in housing units from 1978 to 2001 and increases in home floorspace, air conditioning, and appliances in those units.[7–11] Technological improvements in home construction and improvements in the efficiency of major energy systems (heating, air conditioning, and water heating), lighting products, and individual appliances (from refrigerators to computers) have helped to offset the increased demand for residential energy services. Also, the housing unit population grew most rapidly in the warmer South and West regions.

In recent years, space heating, the largest single home energy requirement, has migrated somewhat out of oil and into electricity. Natural gas retains its majority status for home heating; about 55% of homes used natural gas as their main space heating fuel in both 1980 and 2001. Similarly, liquefied petroleum gas (LPG) keeps its 4%–5% share in both years. But in 1980, electricity was neck and neck with oil as a main heating fuel (an estimated 18.5 and 17.4% of households, respectively, used them for main space heat). By 2001, electricity was the main space heating fuel in almost 30% of housing units, and oil provided the main space heat for only 8.3% of units. This shift reflects the migration of the population and the booming construction in the South and West regions of the country, where lesser heating requirements make electric heat more economically feasible, and the substantial additional entry of natural gas into the Northeast, formerly an area much more dependent on fuel oil for heating.[12,13]

Electric air conditioning is commonly thought to be a major contributor to household energy use, but its importance is not as great as one might expect. In 2001, an estimated 16% of electricity use was devoted to air conditioning, whereas four times as much household electricity use goes to lighting and appliances.[14] On the surface, this result may seem rather surprising. However, appliance use occurs in virtually all homes, whereas air conditioning is absent from almost one-fourth of households, and appliances are used year round, whereas air conditioning is only a few-month activity in many parts of the country.

Given the proliferation of appliances in homes in recent years, it may also seem surprising that appliance electricity use has not gained significant share over the years. However, some newer appliances are specialty gadgets that use relatively little energy, and others are somewhat redundant (the third and fourth televisions), so that use of each individual appliance goes down.

Commercial

The large majority of energy use in commercial, or service, facilities takes place in commercial buildings, defined here as structures totally enclosed by walls from the foundation to the roofline, intended for human occupancy, with a majority of floorspace devoted to commercial/service activity. A wide and interesting variety of service activity is associated with nonbuilding facilities, such as open parking garages, transmission towers, drawbridges, billboards, lumber yards, street lights, parks, open amphitheatres, outdoor sports facilities, and docks and piers. However, all this activity comprises a relatively small share of total commercial energy demand, based on comparisons between EIA's supply side data programs, which account for demand by all commercial customers, and its consumer survey, the Commercial Buildings Energy Consumption Survey (CBECS).[15–17] It should be noted that no known national survey of nonbuilding commercial facilities has ever been undertaken.

Latest building data from the CBECS, for 2003, indicates that there were slightly fewer than 5 million commercial buildings in the United States of 1000 ft^2 or larger. The building stock encloses about 71 billion square feet (more than 2500 mi^2) of floorspace, an area almost as large as the combined land area of Rhode Island and Delaware. Four types of buildings dominate the stock: education, mercantile (sales), office, and warehouse/storage. Together, these four types comprise about one-half of all commercial buildings and 60% of the floorspace. Additional major classes of activity in commercial buildings include assembly (e.g., auditoriums, theatres, civic and convention centers), health care, food sales, food service, religious worship, and several others.[18]

However, space dominance does not necessarily translate into energy dominance. Fig. 2 shows the shares of building stock, floorspace, and major energy consumption in 1999 for selected major building activities identified in the CBECS. Notable variations among the shares include the much larger shares of energy use than floorspace associated with three very energy-intensive types of buildings: health care (especially hospitals), food service (restaurants/cafeterias), and food sales (grocery stores). Each of these activities has its own set of reasons for its high energy requirements: health care due to space conditioning, water heating, and its sophisticated equipment; food service due largely to cooking and water heating, as well as refrigeration of stored food; and food sales due to space conditioning and refrigeration requirements.[19]

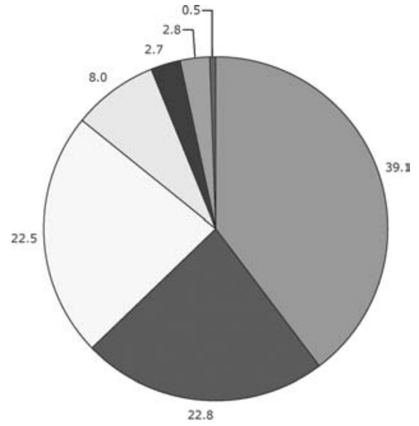


Fig. 2 Distriution of energy use by type of energy: United States, 2003.
Source: From Energy Information Administration, Annual Energy Review 2003. Table 1.3 at http://www.eia.doe.gov/emeu/aer/pdf/pages/sec1_9.pdf (accessed May 2005).

Industrial

Of the major industrial activities mentioned earlier, only manufacturing is covered by an EIA consumer-based survey, the Manufacturing Energy Consumption Survey (MECS). Manufacturing has historically accounted for the great majority of industrial energy use, based on comparison with EIA supplier survey information.[20,21] Although much has been made recently of the departure of substantial amounts of manufacturing from the United States, the energy use, including energy embodied in electricity, associated with our manufacturing alone is still enough to rank the United States second in total energy use among the world's other nations, behind only China.[1,20]

Manufacturing is an especially interesting sector in which to examine energy use, because the throughput of energy at manufacturing plants is much more complex than it is in most other parts of the economy. Most other energy users obtain outside energy sources and use them to provide heat and power, with no further energy accounting required. Manufacturers obtain large quantities of outside energy sources, but sometimes also produce their own energy from sources such as hydro, onsite mines, and wells. However the energy arrives, much of it is used for heat and power, but some applications use energy sources for their chemical or physical content, as so-called feedstocks or raw material inputs to manufacturing processes that result in nonenergy products. One prominent example is the use of natural gas in producing ammonia and fertilizers. Another is the enormous petroleum inputs used to produce product outputs not subsequently treated as energy: asphalt and road oil, lubricating oils, wax and grease, and certain petrochemical feedstocks. Also, manufacturing processes produce byproduct streams, such as blast furnace gas at steel mills, still gas at petroleum refineries, wood residue at pulp and paper mills and plant residues at food processing plants—all can be reclaimed to provide heat or power. Manufacturers routinely transform energy inputs onsite, whether traditional or byproduct, into different forms, primarily electricity, steam, and hot and chilled water, for later use in manufacturing or managing the facility environment.[22]

In 2002, the latest year for which detailed MECS data are available, total manufacturing energy use for all purposes, with electricity counted at its site energy value, was 22.7 quads. Of this, 12.0 quads, or 53%, were accounted for by traditional energy inputs obtained from offsite and used for heat and power requirements. The remainder was onsite, byproduct, and waste product energy, and energy used as feedstock. Fig. 3 shows that 6 of the 21 major manufacturing industry groups account for 88% of total manufacturing energy use: food, paper, petroleum and coal products, chemicals, nonmetallic minerals (stone/clay/glass), and primary metals. These industries involve complex processes that require enormous amounts of heat and power to break down and reform primary materials, and substantial use of energy as raw material inputs. Other later-stage industries, such as assembly of vehicles, computers, clothing, appliances, and furniture, are much more intricate and are certainly valuable in their contribution to the country's economic output, but their energy requirements are much smaller than those of the dominant industries. A large portion of the energy inputs in manufacturing go into boilers for producing heat and electricity, power machinery, and direct process heat applications, such as kilns, heat treaters, and furnaces.[23]

Direct data on energy use for the other industrial applications—agriculture, forestry, fisheries, mining, and construction—are much more scattered, and in some cases not available. The EIA assembles energy use data for nonmanufacturing industries as part of its efforts to model future energy use. Data from this effort, assembled from the

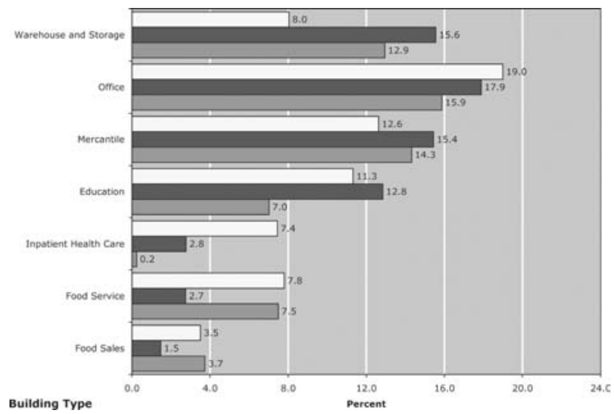


Fig. 3 Percent of commercial building stock, floorspace, and energy use for selected building tyes: U.S., 1999.
Source: From Energy Information Administration (see Ref. 16).

Census Bureau's Economic Censuses and other data sources, has agricultural energy use at about 1.2 quads in 2002, mining use at about 2.5 quads, and construction use at about 0.9 quads.[24]

Transportation

The movement of people and goods is a vital part of the U.S. economy and society. Transportation accounts for about one-fourth of U.S. energy use, but it is an enormously important fourth, because it is virtually all petroleum.[25] In fact, transportation accounts for about two-thirds of petroleum use in the United States. The sector consists of highway vehicles, including cars, trucks, motorcycles, buses, and other specialty vehicles; rail transport, both local and intercity/long distance; aircraft; marine craft, such as ships, smaller boats, and barges; and pipelines carrying a variety of commodities. These classes of vehicles use various grades of petroleum products: gasoline, light diesel, heavier diesel, jet fuel, and heavier, residual type bunker oil (for ships). The largest share of non-petroleum energy used for transportation is for natural gas pipeline transport.[26]

Transportation can be a difficult sector to categorize and study. It is enmeshed with all the other end use sectors, because they all use transportation or transportation services to carry out their roles. There are firms dedicated to transportation service (i.e., movers and other for-hire shipping companies, taxi companies, rental car companies, airports, cruise and ferry lines, towing services, etc.) whose holdings include buildings that are categorized as commercial, but whose vehicles are unquestionably characterized as transportation. Firms and organizations whose primary activities are not transportation service use vehicles for pizza delivery, landscaping service, funeral transport, transportation to school, traffic or weather observation, crop dusting, and so on, in vast array, in the course of carrying out their primary function.

Vehicle use can thus be split into terms of residential vs nonresidential, passenger vs freight, highway vs nonhighway, and by type of vehicle or transport medium. And to add more confusion, some vehicle use is commonly thought of as non-transport! For the purposes of categorizing energy use, some vehicles are commonly considered part of the sectoral activity they are supporting, rather than part of a separate transportation activity. Examples include forklifts at the local furniture store, hardware store, or warehouse; bulldozers and earthmovers at construction sites; captive vehicles used at commercial or industrial sites, but not registered for on-road use; vehicle rides at commercial theme parks and amusement parks; farm implements; and our riding lawn tractors at home.

The primary vehicles in our society are our household vehicles, which enable us to lead our lives in a way that could not even have been imagined a century ago. These

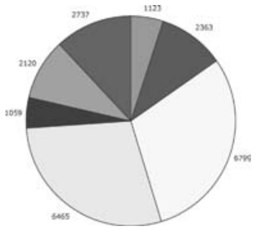


Fig. 4 Distribution of energy use by major manufacturing industry groups: United States, 2002 (Trillion Btu).
Source: From Energy Information Administration, 2002 Manufacturing Energy Consumption Survey. Table 1.2 at http://www.eia.doe.gov/emeu/mecs/mecs2002/sheltables.html (accessed May 2005).

vehicles, about 190 million of them, account for slightly over one-half of all transportation energy use, and about 35% of total U.S. petroleum use. Household vehicles used to be primarily automobiles, with a relative scattering of pickup trucks and "Jeeps." However, since 1980, pickup trucks have proliferated, Jeeps have evolved into our present large class of Sport Utility Vehicles (SUVs), and a new class of household vehicle, the minivan, has evolved from the old business van. As a result, over 50% of current household vehicle sales are one of these non-automobile categories, which now make up about 40% of the household vehicle stock.[27]

Pickup trucks, vans, and SUVs generally have somewhat lower fuel economy than their automobile counterparts. Therefore, their increasing presence in the vehicle stock has been one of the factors (along with increased horsepower and vehicle weight) offsetting engine, equipment, and vehicle design technology, all of which tend to increase fuel economy. As a result, the fuel economy of new vehicles has been relatively constant since the late 1980s (Fig. 4), though the overall fuel economy of the vehicle stock continued to rise for years after that as newer, better-mileage vehicles augmented the stock and replaced the vast majority of older, less technologically advanced vehicles.

CONCLUSION

Obviously, the workings of the U.S. economy, and indeed our entire society, are linked inextricably to energy consumption. Our geographical size, wide variety of climate, large population relative to that of most other nations, economic wealth, and high levels of personal and organizational activity combine to create enormous demands for energy. These demands continue to exert pressure on energy supplies and the infrastructure for energy delivery. The United States already makes use of a wide variety of energy forms to meet its needs. It is likely that, in the future, we will have to increase our use of unconventional energy that currently accounts for only a small part of our consumption, and bring new types of energy online in order to continue to meet the country's energy requirements reliably and at reasonable cost.

Additional information about energy throughput in the United States can be accessed through the Web site of the EIA at http://www.eia.doe.gov. More details on EIA's consumer surveys and the relationship between consumer characteristics and energy use can be accessed through the Energy Consumption page, http://www.eia.doe.gov/emeu/consumption/.

ACKNOWLEDGMENTS

Many thanks to Mark Schipper of the EIA, who provided invaluable assistance in formatting and manipulating the figures in this entry.

REFERENCES

1. *International Energy Annual 2002*. Energy Information Administration, Washington, DC, 2004. Table E.1 at http://www.eia.doe.gov/pub/international/iealf/tablee1.xls (accessed March 2005). A British thermal unit is the amount of energy required to raise the temperature of one pound of water from 39 to 40 degrees F, about the amount of energy provided by a kitchen match. A gallon of motor gasoline contains about 125,000 Btu of thermal energy, and a typical home natural gas furnace is rated at 100,000 Btu used per hour of operation. A quadrillion Btu is a billion million Btu, enough energy to run that natural gas furnace continuously for more than a million years, or to drive around the World 6.4 million times in a 20 mile-per-gallon vehicle.
2. *International Energy Annual 2002*. Energy Information Administration, Washington, DC, 2004. Table E.1c at http://www.eia.doe.gov/pub/international/iealf/tablee1c.xls and Table E.1g at http://www.eia.doe.gov/pub/international/iealf/tablee1g.xls (accessed March 2005).
3. *March 2005 Monthly Energy Review*. Energy Information Administration, Washington, DC, 2005. Table A6 at http://tonto.eia.doe.gov/merquery/mer_data.asp?table=TA6 (accessed April 2005).
4. http://www.genesisny.net/Transport/Hoppers.html (accessed April 2005).
5. *January 2005 Electric Power Monthly*. Energy Information Administration, Washington, DC, 2005. Table 1.1 at http://www.eia.doe.gov/cneaf/electricity/epm/table1_1.html (accessed April 2005).
6. *The Restructuring of the Electric Power Industry: A Capsule of Issues and Events*. Energy Information Administration, Washington, DC, January 2000, p. 8, at http://tonto.eia.doe.gov/FTPROOT/other/booklet.pdf (accessed April 2005).
7. *Changes in Energy Usage in Residential Housing Units*. Energy Information Administration, Washington, DC, 2000, at http://www.eia.doe.gov/emeu/recs/recs97/decade.html#geolocat (accessed May 2005).
8. *The Effect of Income on Appliances in U.S. Households*. Energy Information Administration, 2003, at http://www.eia.doe.gov/emeu/recs/appliances/appliances.html (accessed May 2005).
9. *Residential Energy Consumption Survey 2001 Consumption and Expenditures Data Tables*. Energy information Administration, 2004, at http://www.eia.doe.gov/emeu/recs/recs2001/detailcetbls.html (accessed May 2005).
10. *Residential Energy Consumption Survey 2001 Housing Characteristics Data Tables*. Energy Information Administration, 2003, at http://www.eia.doe.gov/emeu/recs/recs2001/detail_tables.html (accessed May 2005).
11. *Square Footage Measurements and Comparisons: Caveat Emptor*. Energy Information Administration, 2003, at http://www.eia.doe.gov/emeu/recs/sqft-measure.html (accessed May 2005).
12. *Residential Energy Consumption Survey: Housing Characteristics, 1980*, Publication DOE/EIA-0314; Energy Information Administration. U.S. Government Printing Office: Washington, DC, 1982; Table 15, 71.
13. *Residential Energy Consumption Survey 2001, Housing Characteristics Data Tables*. Energy Information Administration, Washington, DC, 2004, at http://www.eia.doe.gov/emeu/recs/recs2001/hc_pdf/spheat/hc3-9a_ne_region2001.pdf; www.eia.doe.gov/emeu/recs/recs2001/hc_pdf/spheat/hc3-11a_so_region2001.pdf; and www.eia.doe.gov/emeu/recs/recs2001/hc_pdf/spheat/hc3-12a_westregion2001.pdf, respectively, for the Northeast, South, and West regions (accessed April 2005).
14. *End Use Consumption of Electricity*. Energy Information Administration, Washington, DC, 2004, at http://www.eia.doe.gov/emeu/recs/recs2001/enduse2001/enduse2001.html (accessed April 2005).
15. *Annual Energy Review 2003*. Energy Information Administration, Washington, DC, 2005, at http://www.eia.doe.gov/emeu/aer/pdf/pages/sec2_6.pdf (accessed April 2005).
16. *1999 Commercial Buildings Energy Consumption Survey Detailed Tables*. Energy Information Administration, Washington, DC. 2002. Table Cl at http://www.eia.doe.gov/emeu/cbecs/pdf/alltables.pdf (accessed April 2005). These energy data are for calendar year 1999. The energy consumption data from the 2003 CBECS have not been released at the time of this writing. Caution needs to be taken in comparing energy use numbers taken from supplier surveys with consumer-based numbers for CBECS. Supplier values are based on customer classes rather than facility activity. Facilities such as large apartment buildings and small manufacturers are often classified by energy suppliers as commercial accounts. Conversely, sometimes a large office complex or other type of institutional campus might be classified as an industrial account.
17. *Preliminary Data: 2003 Commercial Buildings Energy Consumption Survey, Commercial Buildings Characteristics*. Energy Information Administration, Washington, DC, 2005, at http://www.eia.doe.gov/emeu/cbecs/cbecs2003/introduction.html (accessed April 2005).
18. *A Look at Commercial Buildings in 1995: Characteristics, Energy Consumption, and Energy Expenditures*, Publication DOE/EIA-0625/95; Energy Information Administration. U.S. Government Printing Office: Washington, DC, 1998; 20–21.
19. *2002 Energy Consumption by Manufacturers—Data Tables*. Energy Information Administration, Washington, DC 2005,

at http://www.eia.doe.gov/emeu/mecs/mecs2002/data02/excel/table1.2_02.xls (accessed May 2005).
20. *Annual Energy Review 2003*. Energy Information Administration, Washington, DC, 2005, at http://www.eia.doe.gov/emeu/aer/pdf/pages/sec2_7.pdf (accessed May 2005).
21. *Manufacturing Consumption of Energy, 1991*, Publication DOE/EIA-0512(91); Energy Information Administration. U.S. Government Printing Office: Washington, DC, 1994; Figure 2.2, 10–11.
22. *End Uses of Fuel Consumption, 1998*. From the Manufacturing Energy Consumption Survey, Energy Information Administration, Washington, DC, 2004. Table N6.2, at http://www.eia.doe.gov/emeu/mecs/mecs98/datatables/d98n6_2.htm (accessed April 2005).
23. *Annual Energy Outlook 2005*. Energy Information Administration, Washington, DC, 2005. Table 32 of supplemental tables at http://www.eia.doe.gov/oiaf/aeo/supplement/sup_rci.xls (accessed April 2005).
24. *Annual Energy Review 2003*. Energy Information Administration, Washington, DC, 2005, at http://www.eia.doe.gov/emeu/aer/pdf/pages/sec2_8.pdf (accessed April 2005).
25. Conversation with John Maples, EIA transportation analyst, April 25, 2005.
26. *Annual Energy Review 2005*. Energy Information Administration, Washington, DC, 2006, Table 2.1e; at http://www.eia.doe.gov/emeu/aer/pdf/pages/sec2_8.pdf (accessed November 2006).
27. *Household Vehicles Energy Use: Latest Data and Trends*. Energy Information Administration, Washington, DC, 2005, at http://www.eia.doe.gov/emeu/rtecs/nhts_survey/2001/index.html (accessed November 2006).

Energy Use: U.S. Transportation

Kevin C. Coates
Transportation & Energy Policy, Coates Consult, Bethesda, Maryland, U.S.A.

Abstract
Oil, gas, and electricity prices are spiking, America's highways and airports are overcrowded, suburbia has morphed into urbania, and suburban sprawl has led to exurban sprawl. As political and industry leaders continue to apply old solutions to solve the problems of rapid growth, they are discovering that those old solutions do not work in the face of exponential growth. The nearly simultaneous convergence of exponential growth factors in the demand for energy and transportation will eventually force America to reassess its present land development policies and find ways to link this new policy with more sustainable energy and transportation solutions. This paper not only looks at the root causes behind America's most pressing national issue, but also offers some workable and sustainable solutions for reversing what is most certainly a worsening situation for most American citizens.

Oil prices continue to climb and continue to reach new heights with each passing year. These increased prices impact the price and availability of chemical products where oil is the feedstock, as well as the cost and availability of oil-dependent transportation fuels, and they will increase the price of most other goods and services.

Soaring U.S. and world demand for oil has another consequence. Highway and road transportation systems are grinding to a halt under the crush of peak-time travelers, with average trip times taking ever longer in most U.S. cities.[1] In addition, airports suffer from "winglock," which leaves fuel-hungry jets idling on tarmacs while they wait for a gate or a runway while arriving flights are forced to slow down to avoid energy-wasteful holding patterns.

There are far deeper issues involved here than simply too many cars or planes in the same place at the same time or the price of a gallon of gasoline. The country's overburdened transportation system is really symptomatic of a more deeply rooted psychological malady—America's energy attitude. It could be that higher energy prices will provide the wake-up call America needs to change its profligate energy ways—not the least of which is the inefficient Helter-Skelter development of cheap land farther and farther from urban centers.

Weaned on the habit of cheap and widely available energy, generations of Americans have been seduced by the idea that they can build or travel wherever, whenever, and however they desire. Indeed, Americans have even come to equate mobility with their sense of liberty.

The real question is, is America truly free if it continues to build and attempt to maintain an entire societal infrastructure built upon the premise of endless supplies of cheap energy, especially when (1) 96.9% of its transportation system is dependent on oil for its fuel derivatives,[2] (2) these oil-based fuels rapidly rise in price and are extremely price volatile, and (3) two-thirds of the oil used to fuel the U.S. transportation system is dependent on foreign nations?

Few Americans know or care how much energy it takes to propel a single driver in a car at 60 mph. Slick advertising campaigns lure customers into buying 265-horsepower machines to roar down empty asphalt tracks in a powerful pursuit of freedom. Few people consider how much energy it takes to build each vehicle or to clear, construct, pave, and endlessly repave hundreds of thousands of miles of oil-based asphalt (Fig. 1).

America's ground transportation and energy problems are the direct result of failed land development policies. Where buildings are constructed directly impacts the development of road networks, influences the energy consumption levels of road trips, and dictates traffic patterns. As automobile ownership soared in the mid-20th century, America became less concerned with building lasting structures and communities scaled to human-powered transportation or activities, and focused more on quick construction and building an infrastructure that focused instead on accommodating its transportation machines. Think about how long it takes to walk between distant gates at Chicago's O'Hare airport vs the compact layout of a big-city train station. Not only does this overriding focus on meeting the needs of machines consume vast quantities of energy, but it also creates environments hostile to pedestrians—that's you and me when not traveling in machines.

The fact that America has three separate and distinct federal departments for housing, transportation, and energy reveals a lack of appreciation for the increasing importance of the energy correlation. In Germany, for

Keywords: Maglev; Energy; Oil; Natural gas; Hubbert; Peak oil; Sprawl; Energy policy; Transportation; Policy.

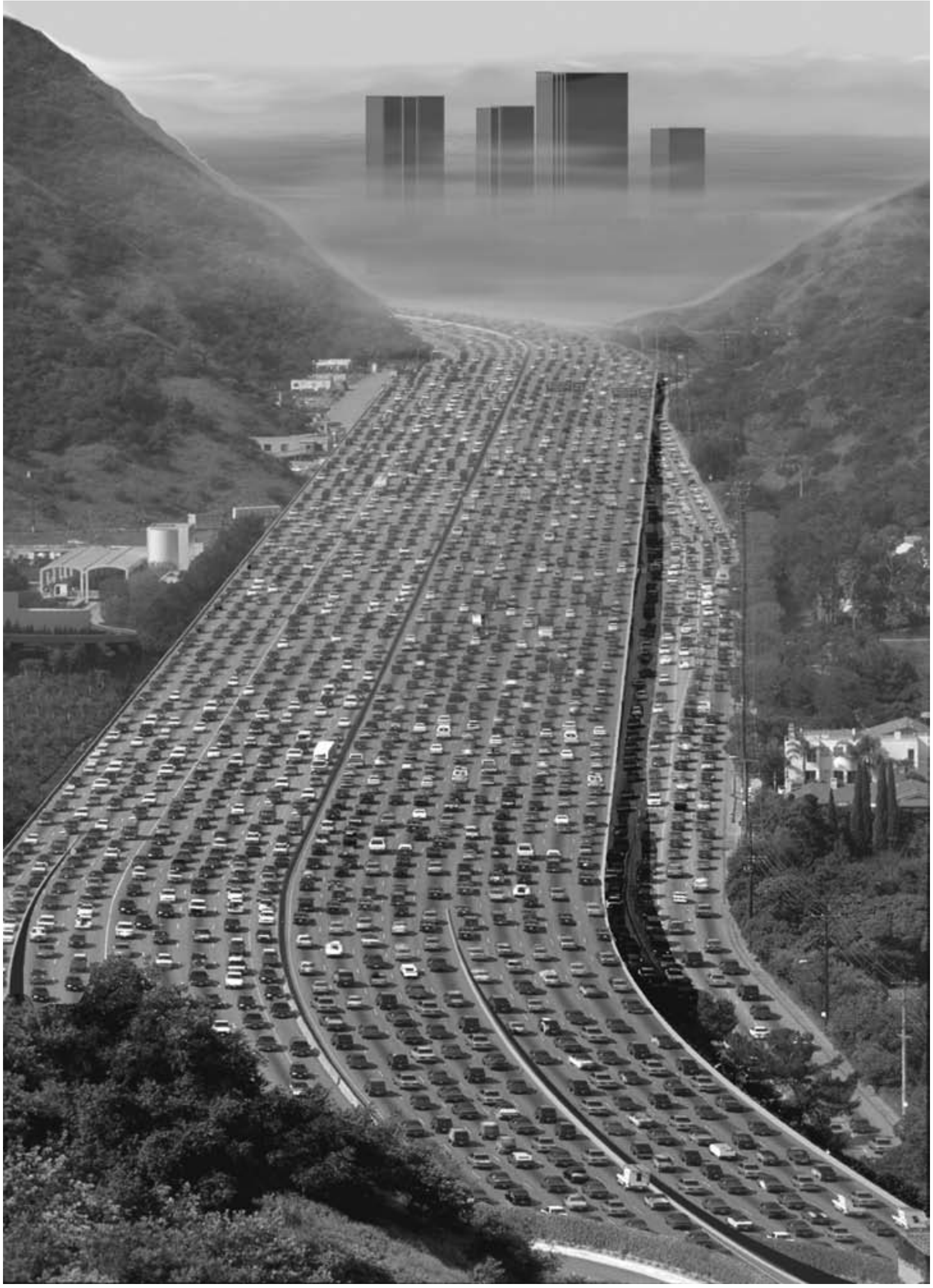


Fig. 1 Is this the future we want? Computer generated view of the 405 north of Los Angeles, courtesy of Phil Shisbey, Lake Elsinore, CA (www.transfuture.net). Note the Getty Museum trams on the left and the smog in the San Fernando Valley.

instance, housing and transportation are combined into one federal ministry, the Bundesministerium für Verkehr, Bauwesen, und Wohnungswesen (Federal Ministry of Transport, Building, and Housing). Japan has its Ministry of Land, Infrastructure, and Transport because land is limited, natural energy resources are scarce, and the efficient movement of people and goods between population centers keeps energy transportation costs low.

Industrialized countries around the world were forced years ago to recognize the "capacity limitations" of their infrastructure and domestic resources, especially their potential domestic sources of energy. In contrast, America's infrastructure was built upon the premise of widely available cheap land, endless supplies of cheap domestic energy, and relatively cheap personal transportation—a system that has nurtured an economic belief in growth without limits. Well, there are limits to exponential growth, and the negative effects are now becoming evident.

The power of an exponential equation is illustrated in the old fable about a ruler in ancient Persia who wanted to reward a subject for having created the game of chess. Rather than requesting gold or jewels, the clever inventor asked instead for one grain of wheat placed on the first of the 64 squares on the chessboard. For each successive square, he asked that the amount of wheat be doubled. Not being a mathematician, the perplexed ruler agreed. Unfortunately, long before reaching the 64th square, all of the wheat in the kingdom had been exhausted. Indeed, there is not enough wheat in the world today to satisfy this formula—nor is there enough oil (Fig 2).

Approximately 65% of American oil consumption is for transportation. The country's two primary transportation modes, road and air travel, are entirely oil dependent. These modes served America well enough (albeit with high pollution levels) when fuel was relatively cheap and the country's population rate of growth was slow, but as already mentioned, the price of all hydrocarbon fuels is rising, driven by a burgeoning worldwide consumer population.

What Americans are beginning to experience is the convergence of higher energy costs and the almost exponential increase of the number of vehicles on its capacity-limited highways. Indeed, the instant replacement of all of America's internal combustion powered vehicles with hybrid or electric cars would not solve its traffic congestion problems. The problem with relying almost exclusively on highways and roads for transportation is that they are inherently "capacity inflexible," meaning that massive amounts of new construction are needed to accommodate each successive doubling of the road-traveling population.[2] What used to be horrendous unidirectional rush hour traffic in the mornings and evenings has now extended into most of the day in both directions throughout most of the country.

As the periods shorten between the doubling of road populations, no amount of road construction will be able to

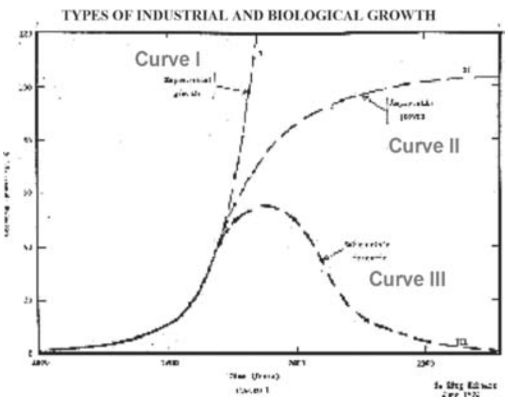


Fig. 2 Types of industrial and biological growth. This is a chart with three types of growth curves. Curve I is for exponential growth. Notice the similarity that Curve I has with the curves in the following two graphs. Curve II reflects growth that plateaus to a sustainable level (not applicable to a finite resource being depleted). This chart was shown during a 1974 Congressional subcommittee hearing on energy as part of M. King Hubbert's testimony; the geophysicist who predicted correctly in 1956 that U.S. oil production would peak in 1970 and then begin to decline (as modeled by Curve III). Hubbert's prediction proved remarkably accurate.

keep up with such demand. Despite ample evidence that proves this strategy is not working, federal, state, and local governments persist in allowing the building of houses farther and farther from urban centers, and hence force all taxpayers to subsidize exurban development by picking up the tab for the new road infrastructure.

The rapidly increasing need for new transportation construction combined with the costly issue of maintaining the vast infrastructure already in place is the major reason so many state highway budgets are in the red. Just to keep pace with the highway construction-intensive transportation system, the American Society of Civil Engineers and the House and Senate transportation subcommittees recommended in 2003 that funding levels for a new federal transportation bill be approximately 70% higher than the $218 billion authorized in the 1998 six-year TEA-21 bill,[3] which itself was 44% higher than the previous Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA).[4] The White House insisted that the two houses of Congress submit a bill to the president that was well below $300 billion or face a veto. For its part, the White House supported a $256 billion, six-year bill that was only $45 billion above the 1998 six-year TEA-21 bill, or only a 21% increase. Clearly, there is a disconnect here between reality and policy, given the exponential growth and demand on the system.

The transportation bill's various critics in the media blamed "pork barrel politics" for the steep funding requests. The reality is that most of the increases are more the direct result of exponential housing growth farther and farther from downtowns and the need for necessary repairs of an aging and overused highway, road, bridge, and tunnel infrastructure. Developer sprawl bleeds funding from existing transportation infrastructure maintenance to feed new road construction, thus creating a funding crisis in some states and forcing some governors and state legislatures to seriously consider increasing state gasoline taxes in spite of the threat of political suicide. Despite these realities, the 2005 transportation bill passed by Congress and signed by the president was predominantly a highway bill, and as a result, it represents another missed opportunity to pursue more workable and sustainable transportation solutions.

Conversely, transit system operators need only to add more cars to trains and run trains more frequently to accommodate periodic bursts in ridership and to avoid disrupting passenger flow. Furthermore, rail- or guideway-based transit infrastructure with dedicated right of ways (ROWs) can sustain longer periods of steady ridership growth (capacity flexible) before new infrastructure construction is deemed necessary.

Unfortunately for today's Americans, when U.S. leaders decided in the mid-1950s that rail travel was old fashioned and of low value, unprecedented levels of federal tax dollars were spent for airport expansion and launched the massive infrastructure development projects that became the present interstate highway system. This 1950s policy decision locked American transportation into today's oil dependence.[5] Massive federal funding for roads and airports not only eroded the many private rail operators' passenger base, but also eliminated rail as a viable alternative mode of transportation for taking ridership pressure off the other two modes. Entire rail networks have disappeared, such as the 1600-mile Pacific Electric Railway system that provided an efficient and affordable mass transit system in the first half of the 20th century for Los Angeles area residents—an area now all too frequently stuck in highway gridlock. With so many of America's roads and highways now suffering the same fate, the oil-dependent transportation policy born in the 1950s is proving to be shortsighted.

Besides being more capacity flexible, rail or guideway-based systems can also more easily use electricity for power because they operate "within and along an electricity corridor or grid." This also means they are not reliant on any one prime energy source and they do not introduce air pollution along their routes. Not only are these systems energy efficient, but they also are more cost effective because their energy costs are mitigated by the several primary energy sources that comprise electricity generation, which is usually attracted to the lowest-cost production methods. In addition, these transportation systems allow for the seamless introduction and integration of clean renewable energy sources at no additional cost. In other words, as new energy sources harnessed from the wind, the waves, the sun, and geothermal sources come online, they can seamlessly be injected into the existing power mix that is propelling 60 or 300-mph transit vehicles across the landscape. All the technology necessary to make this a reality is now available and in operation in various locations around the globe (Fig. 3).

By meshing the electricity industry's new power technologies into a cohesive network that is combined with a new transportation paradigm of high- and low-speed transit, such as magnetic levitation (maglev) systems, the country can create an extremely robust "electricity transportation grid" that not only upgrades America's aging electricity grid, but also revitalizes its overburdened transportation system. The combined result would introduce higher levels of reliability for both systems and improve trip times, reliability, safety, and comfort for intercity travel, while enabling and facilitating the wholesale transmission and distribution of electricity on a national scale (something for which the present grid was never designed). This combined approach would also cost less than pursuing either system upgrade independently because both would share the same ROW acquisition costs. Indeed, it is the difficulty of securing new ROWs for electricity transmission that has significantly hampered expansion and the reliability of America's power grid. By enclosing new transmission cables in conduits along a maglev ROW, not only would the site of transmission towers and power lines be eliminated, but also, disruption to electrical power flow resulting from inclement weather would be drastically reduced, resulting in higher electric transmission reliability, improved power flow capacity, and lower overall maintenance costs, which in turn should keep electricity rates low and stable.

One example of a capacity limit that is impacting the price of fuel in America is the international demand for oil tankers, which in late October 2004 exceeded the number of available ships. The countries that build oil tankers, such as Japan and Korea, are at capacity and building them

Fig. 3 Elevated 267 mph Shanghai Maglev is on time 99.9% of the time and places a minimal footprint on the landscape and does not cut communities in half or obstruct their existing road networks. Photo courtesy of Transrapid International.

as fast as they can. Additionally, it takes four years from the time an oil tanker is ordered until it is delivered into service, and the number of new tankers coming online in 2005 is not likely to keep pace with the projected increase in world demand. Exacerbating this already dicey shipping supply situation was an international maritime agreement that mandated the scrapping of all tankers built before 1982 by May 2005 and the scrapping of all non-double-hulled tankers by 2010. In October 2004, shipping rates for a load of crude jumped from $35,000 per day to $135,000 because of the shipping shortage, meaning that the cost of a typical 40-day cruise jumped $4 million.[6] This translated into a gasoline price increase of about a nickel per gallon for motorists and a further reduction of discretionary spending for other goods and services. Limited world refinery capacity is another choke point that is further constricted by temporary scheduled and unscheduled maintenance shutdowns and by permanent shutdowns caused by political pressures to lower pollution levels.

As the price of oil crept over the $50-per-barrel mark in early March 2005 and as traffic congestion continued unabated, there was little media attention. When the price hit $55 per barrel by midmonth and kept going, the media and the public finally took notice. Make no mistake about it—this is a serious omen for America. The problem is not that the world is running out of oil anytime soon, but that demand is exceeding the capacity limits of infrastructure to deliver and refine the oil being extracted from wells. Compounding America's energy problem is the fact that Asians hold over $2 trillion worth of its debt and are making noises about divesting themselves of dollars in exchange for euros or yen.[7] If oil exporting nations begin trading oil for euros instead of dollars, such a move would make oil even more expensive for Americans or even result in denied access to some oil supplies, thus setting up shortage scenarios or even oil wars.

Remedying these capacity limitations will cost billions of dollars and take years to implement. So why didn't all of the industries associated with the oil industry, such as shippers, refiners, and economic analysts, anticipate the rapid increase in oil demand years ago?

The main cause for seemingly sudden capacity problems is that most government and private-sector energy and economic experts miscalculated the precise impact that China's and India's rapid economic growth would have—not only on the world's supply of oil, but also on all commodities and available shipping. For instance, the Energy Information Agency (EIA) predicted in its May 2003 International World Outlook Report a world increase of 3.1% per year until 2025—far above the long-term annual trend rate of 1.4% that occurred through 1989–1999.[8] These EIA predictions were widely reported, especially the part that world oil prices would remain around $25 per barrel until a production peak of 120 million barrels per day (Mbd) was reached in 2020, even with a projected 60% increase in global energy demand and steady economic growth in developing nations.[9] How could they have been so far off? (Fig. 4)

One reason for confusion among experts is that China's political and economic system is not transparent, making accurate predictions of the demand for oil and other commodities impossible.[10] Further exacerbating the world oil economics prediction game is the lack of transparent data from the world's leading oil exporting countries for evaluating their maximum oil production capabilities.[11] This "oil reserve uncertainty" amounts to a huge accountability crisis for world economies and has the potential for causing a worldwide depression. Without knowing the actual maximum rate of oil production, governments, energy companies, financial markets, and other professionals have had no accurate way to evaluate the breadth and depth of supply. In spite of uncertainty in both supply and demand, demand remains high. Demand will remain high and even go higher until distribution bottlenecks force higher prices for oil and gasoline, eventually leading to a stifling of world economies until some sort of balance is restored between supply and demand.

Besides conventional oil supplies being severely restricted, oil production is now extremely vulnerable to a single contingency disruption. The world derives fully 20% of its oil from the 14 largest fields, each with an average age of 40-plus years.[11] Hundreds of medium-size and thousands of small fields produce the world's balance. It is well known in the oil industry that the largest oil fields have a tendency to be discovered first. And whatever an oil field's size, once tapped, it will run its production course and inevitably peak, as have the supergiant Prudhoe Bay, North Sea, and Texas oil fields. History has also demonstrated that once a field peaks, its production falls off sharply. Saudi Arabia's Ghawar oil field, by far the world's largest, might produce oil for another hundred

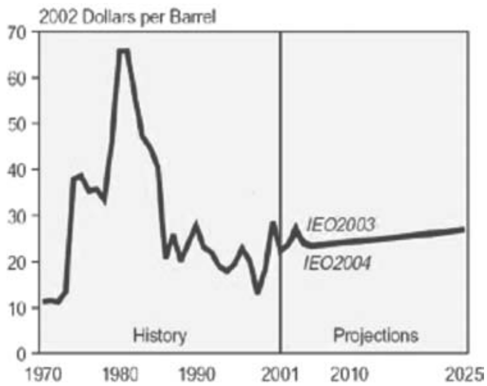


Fig. 4 Comparison of 2003 and 2004 World Oil Price projection, 1970–2025, The now infamously inaccurate 2002 EIA projections for world oil prices to steadily rise to $28 per barrel by 2025.

years, but with unknown and almost certainly lower flow rates. It may have peaked already, but no one knows for sure because the data is not publicly available. And because governments can use oil reserve estimates to secure large international bank loans, the temptation for some governments to inflate their reserve estimates is often too great to resist. Such vagaries make open accounting from the Middle East improbable at best (Fig. 5).

Oil-economy optimists tend to be economists, not geologists. Proponents of maintaining an oil-based economy are fond of trumpeting how advances in technology have enabled better yields from existing wells where perhaps as much as 70% of oil once remained underground and unrecoverable. To be sure, those percentages have improved to about 50% with today's technology. However, new technologies and more intensive extraction efforts can also lead to higher production costs. What the optimists studiously avoid mentioning is that new conventional oil field discoveries are averaging much lower yields. In 2003, Matthew Simmons, president of the energy industry's largest investment banking firm, Simmons & Company International, noted that "it could require more than 3000 new oilfields to be found and developed over the next ten years, compared to slightly more than 400 named new oilfields that were discovered in the 1990's—only 2.5% of which had yields in excess of 100,000 barrels per day."[11]

In other words, as world demand continues to increase, newly discovered sources of supply are smaller, more difficult to reach, have higher production costs, and are just barely keeping up with declines in production of existing fields, and not keeping up with increased world demand.

An alarming and related development on January 9, 2004, was the report that Shell Oil had revised its estimates of known reserves downward by 20%. On February 24, 2004, a Center for Strategic and International Studies (CSIS) forum in Washington, District of Columbia. exposed some professional doubts as to the ability of Saudi Arabia to meet future increases in demand. Attending oil executives and government officials from the United States and Saudi Arabia admitted that Saudi capacity would probably stall near current levels and potentially create a significant gap in the global energy supply[12]—a gap with enormous consequences for world transportation systems.

Simmons, who also presented at this CSIS forum, stated that "we need to begin creating a new form of energy to replace some portion of oil and gas use" and that "getting to this new era could take a long time, it all may not work, and the time to begin is today."[13] He went on to point out that Canadian and U.S. natural gas production has peaked, yet "nobody is talking about it." Well, people are starting to take notice since natural gas prices rose to over $14 per million Btu after Hurricane Katrina hit the Gulf Coast.

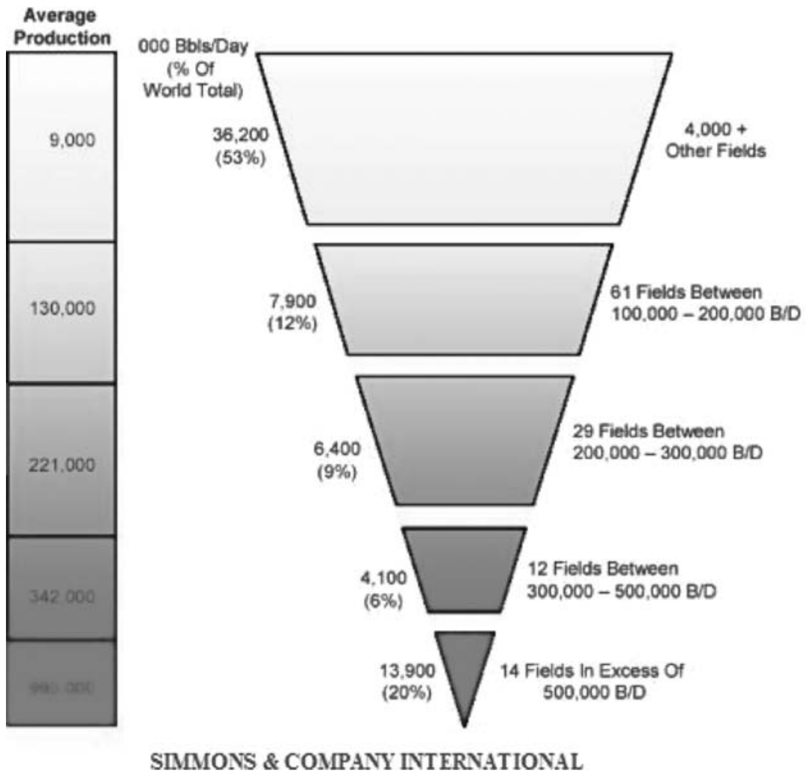


Fig. 5 The oil pyramid.

It just so happens that America's neighbor to the north has tar sand deposits with enough oil trapped in the sands to exceed the known reserves in the Saudi peninsula, albeit not as easy, clean, or cheap to recover. Syncrude, a consortium of eight U.S. and Canadian oil companies, has been actively improving methods for extracting crude from these sands since 1978. In 2003, Shell and Chevron Texaco jointly opened the $5.7 billion Athabasca Oil Sands Project in Alberta (about 250 miles north of Edmonton), which pumps out about 155,000 barrels per day. With oil prices well above the $50-per-barrel level, tar sand extraction can remain profitable and may justify further expansion.

In a well timed report and accompanying press release, Cambridge Energy Research Associates (CERA) stated that as a result of new recovery methods, tar sands represented a vast new oil reserve for the United States.[14] Maybe so, but this oil locked up in tar sands is a mixed blessing for both Canadians and Americans; America can continue its oil addiction, and Canada will be forced to deal with limiting the large amounts of carbon dioxide, sulfur dioxide, and other toxic emissions created from tar sand production.

However, from an energy expenditure standpoint, extracting oil from tar sands may end up not making very much sense, even if it does make monetary sense presently. The Canadian environmentalist Gary Gallon wrote in *The Gallon Environmental Newsletter* in October 2002, "Much of the energy used to extract tar sands oil comes from using natural gas. The natural gas price is closely tied to the price of conventional oil, on a British thermal unit basis. So when the price of oil zooms up, the price of natural gas rises accordingly. In fact, today, it is possible that the price of natural gas could increase more than conventional oil due to its environmental benefits over burning coal for generating electricity. This would leave tar sands production in an even worse economic situation."

Geologist Colin Campbell has pointed out that the Athabasca tar sands oil production uses "stranded gas"—natural gas from nearby fields that is of insufficient quantity to justify an interconnection with the North American gas pipeline grid. Stranded gas is cheap because there are no other customers for it beside the oil sand production companies. "It is running out fast," says Campbell, "and no one is going to use ordinary oil and gas to produce from the tar sands because they are too valuable." If Campbell is right, and tar sands end up being processed with commercially priced natural gas, when all the accumulated energy is tabulated for the creation of a barrel of tar sand oil, the net gain will be minimal, perhaps even a net negative. When the financial and health impacts of the resulting pollutants are factored in, the wisdom of pursuing this energy source becomes highly questionable.

The scenic region where Utah, Wyoming, and Colorado come together is known as the Green River Basin. It is also the site of the largest oil shale deposit in North America, containing more oil than all the known world reserves of conventional crude. Oil shale is supposedly another huge untapped energy source. Oil shale is neither oil nor shale. It is "source rock" that never reached the depths of the "oil window"—the depths between 7500 and 15,000 ft where organic material is transformed into oil. By exposing this source rock to high temperatures, however, the organic material contained within can be "cracked" to form oil.[15] To create the heat to crack oil takes a lot of energy; therefore, oil extracted this way comes at a cost higher than oil extracted via present conventional oil extraction methods or extraction from tar sands. Additionally, once these source rocks are heated, and the organic material contained within is converted into oil and removed, the source rock increases to about double its original volume, making tailings disposal a real financial, physical, and environmental challenge. Besides the expense and technical difficulty of extracting this oil, the damage that would be done to the land could well be far beyond what society is willing to pay for maintaining its car culture. Then again, maybe not…

These oil energy uncertainties not only portend extreme price volatility, but also raise the specter of an overly oil-dependent America once again having to come to grips with shortages. This time, however, any shortages will not be the result of a politically motivated embargo, as in the 1970s, but the result of physical limitations in oil production and distribution capacity. When and how long these periods of shortages occur depends more on what America does nationally to reduce its oil consumption levels significantly (demand) than on what the oil industry does to increase its production or distribution capacity (supply).

The 1970s oil shortages were a warning that the Nixon, Ford, and Carter administrations clearly understood and attempted to act upon with conservation measures. However, the majority of the American people did not want to hear about behaving in an energy-responsible manner. In 1980, Ronald Reagan was elected over Jimmy Carter in no small part because of his platform boasting that there was plenty of oil left in the world and that the conservationist measures of the Carter administration were another example of big government intruding on people's lives. The great disservice done to the American people by this unfounded political optimism was that it spurred increased housing development farther from city centers, put greater demands on highway and road infrastructure, and encouraged the sale of sport-utility vehicles (SUVs) with more powerful, fuel-hungry engines. The Reagan supporters discredited and ridiculed the notions of environmentalists and conservationists, and embraced the idea that increased oil consumption was a sign of American prosperity. This was a policy tailored for big business.

Unfortunately for Americans, conservationist measures are at complete odds with the economic goals of any profit-driven energy provider corporation; the production goals of any major manufacturer of internal combustion engine vehicles; the advertising industry, which thrives on producing automobile television commercials; and the political powers in Washington who have no trouble accepting political campaign contributions from all those who, understandably, do not want to get off the "gravy train." Ironically, even those who profit from the status quo are stuck in the same box as everyone else.

Ultimately, however, future generations will come to view the present extraordinary search for fossil fuels in much the same way that we now view our ancient ancestors' single-minded quest for wood—a quest that led, and is still leading, to the destruction of forests and denuded landscapes. With some enlightened conservationist measures, those forest resources could have been preserved and their use prolonged. If America cut back on government subsidies for oil exploration/production and highway and airport construction, and began instead to subsidize heavily the development of alternative energy regimes and electric-powered transportation systems, larger amounts of precious oil resources could be preserved for future generations. Oil is also the feedstock for plastics and other products, and its fuel derivatives are used in those applications that must have a compact, portable power source: aircraft, emergency vehicles, farm and construction equipment, and assorted military machines.

Long the world's largest consumer of oil, America presently consumes over 20 Mbd, 12 of which is imported and, at $55 per barrel, bleeds $240.9 billion from the U.S. economy per year (data as of late 2005). The next-largest oil consumer and fastest-growth market is China, which averaged 6.5 Mbd in 2004[16] and was hitting peaks of 8.6 Mbd by late 2005. This is out of a world total at the end of 2005 of 83.6 Mbd (1.5 Mbd more than predicted a year earlier).[17] This reveals a staggering rate of growth in world demand that presents enormous logistical challenges for expanding the existing oil extraction, delivery, and refining infrastructure. And it is this rapid worldwide demand that is driving up oil and gasoline prices in America (Fig. 6).

Rising oil and gas prices inevitably result in a plea for politicians to do something. So both sides of the aisle respond by demanding investigations. In mid-2005, many Democrats naïvely asked that President Bush stop refilling the country's strategic oil reserve to provide a short-term reduction in price pressures for something that is really a systemic long-term problem. And for their part, some Republicans used rising oil prices to gain support for opening the Arctic National Wildlife Reserve (ANWR) to oil and gas exploration, while they also quietly opened up nearly 45 million acres of previously protected federal lands in 12 Western states. At $50 per barrel or higher, even a small oil field deposit can be a profitable venture.

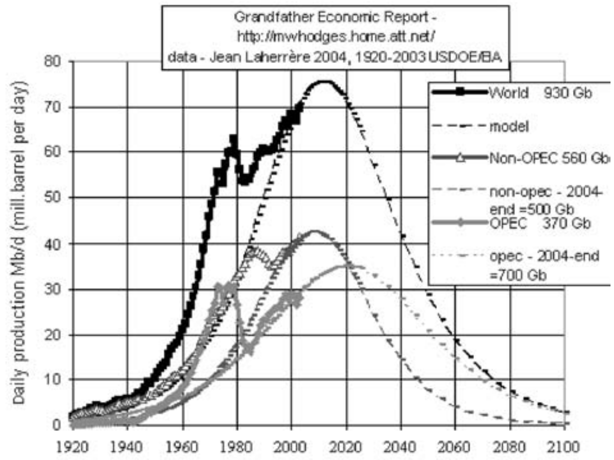


Fig. 6 World conventional crude oil production and forecast for an ultimate of 2.1 Tb (no demand constraint).

Unfortunately, the Republican attempt to drill the country's way out of its energy fix is also naïve because oil supplies can never hope to satisfy what is really a long-term worldwide demand situation.

What is surprising about the ANWR controversy is the lack of any national discussion about who should profit, and by how much, from its oil. After all, the resources to be extracted from ANWR—or any federal land, for that matter—technically belong to the people of the United States. Or do they? According to the General Accounting Office, oil exploration and production (E&P) companies pay either $12\frac{1}{2}\%$ or $16\frac{2}{3}\%$ royalties (a gentler term than taxes) for every barrel produced, depending on whether the oil field is onshore or offshore, respectively.

Practically every other oil-producing nation in the world has pried loose the financial grip of the international oil companies from their natural gas and oil resources: Mexico, Venezuela, Kuwait, Saudi Arabia, Iran, Libya, etc. For years, the oil companies gave these countries a small percentage of the revenue stream until their political leaders caught wise. One by one, these countries began to demand a greater share of the oil profits until they either took complete control of operations or allowed a joint operation partnership that left the oil companies a profit of about 20% or less. The oil companies stayed, figuring that 20% of something was better than 100% of nothing.

To be absolutely clear about what is at stake with ANWR—besides the possible trashing of a small bit of land that few people will ever see—what is at stake is an estimated 10 billion barrels of recoverable oil. At $70 dollars per barrel, that's about $700 billion. Let's just say that by the time this oil comes online in ten years, prices are $100 per barrel. That's over $1 trillion worth of oil. But how will the revenue be divided? According to present law, the government will get only 12.5% of the royalty. But which government gets the royalty? Under the provisions of the

Alaska Statehood Act of 1959, the federal government would receive only 10% of that royalty percentage (only 1.25% of total revenue), with Alaska receiving the other 90%. One has to believe that the majority of Americans would find this outrageous, especially because all Alaskans already receive yearly $2000 royalty checks and don't pay any state income tax. It is no coincidence that Alaskans and Saudi Arabians have similar attitudes toward energy resource ownership and management.

As oil is an internationally traded commodity, oil prices are determined by world markets. The best way for ANWR's oil reserves to help the American people significantly is for the government to take 80% of the revenue at the wellhead. At a production rate of 1 Mbd, this would increase yearly U.S. Treasury revenue from $456 million per year under current rules to $29.2 billion per year for the next 30 years. If Congress doesn't change the royalty system, the American people will get stuck with the triple whammy of basically giving away the oil to E&P companies, paying ever-rising prices to fill their tanks, and then being taxed for fuel at the pump. This is not a Republican vs Democrat issue or an environmentalist vs polluter issue—this is about capturing maximum revenues for the U.S. Treasury to fund properly the rebuilding of America's national transportation infrastructure to be non-oil-reliant for younger and future generations. And most important, it is about doing the right thing.

Given America's mammoth debt and pressing financial needs for Social Security, Medicare, Medicaid, clean and safe shelter for the mentally ill, electricity infrastructure development, and building new alternative electric-powered transportation networks, why is the United States not already charging 80% royalties for the oil under all federal land to fund national necessities? In fiscal 2003, the federal government collected $5.6 billion in royalties from oil and gas production on federal lands, whereas the oil companies presumably collected somewhere around $40 billion. Should the bulk of this money derived from public lands help the oil business earn record profits or should the ratios be reversed, and should the bulk of the money end up in the U.S. Treasury? Maybe this issue should be put in front of the American public as a national referendum.

If every other oil-producing country in the world took back control of its national resources and oil revenue, why doesn't America? Few Americans oppose the idea of a fair profit for E&P companies or allowing them to recoup their admittedly hefty expenses. E&P companies already receive tax breaks and subsidies for their operations, such as deductions for oil transport expenses. And with advances in seismic 3-D modeling for identifying oil deposits and directional drilling technology that allows access into many oil fields from one wellhead, costs have been dramatically lowered, and the likelihood of hitting "wet holes" has increased.

Both parties have declared their desire to lower America's reliance on foreign oil. But the truth is, oil is an internationally traded commodity that is sold to the highest bidder. In spite of laws to the contrary, it is highly conceivable that when ANWR oil production comes online, rather than just transporting it to America, oil companies will also be shipping it to Asia—just as they did with oil from Prudhoe Bay. It is simply too difficult to monitor and trace oil shipments. Regardless of ANWR, Americans will still be using and relying on foreign oil and paying the world's going rate.

America's problem is not reliance on foreign oil, but reliance on oil itself—period—regardless of the source. Because of the very real capacity limitations of world oil production, inevitable shortages will arise in the face of soaring worldwide demand. It is simply a matter of time. The price of oil will rise to unimagined levels as worldwide demand is not met. Then, as demand falls off due to oil no longer being affordable for increasing segments of consumers, prices will recede somewhat and then continue their inexorable upward climb until oil is no longer the predominant and common source of energy for general mass transportation. This price volatility cycle will eventually become a constant companion to all oil-based economies and could also be a major destabilizing ingredient for world democracies. The ramifications are huge.

On March 6, 2005, OPEC's acting secretary general, Adnan Shihab-Eldin, stated that he thought the price would reach $80 per barrel within two years. In a March 2004 interview with geologist Dr Kenneth Deffeyes, he intimated that conventional oil production would peak by November 2005 and that oil could easily hit $80 per barrel not long after. Picking who is right is irrelevant. The point is that America's transportation system simply cannot react fast enough to deflect the negative impacts of such a dramatic upward shift in the pricing levels of its primary transportation fuel (Fig. 7).

In spite of the imminent threat of dramatically rising oil prices and worsening traffic congestion, America's 2005 energy bill fails miserably to address this looming crisis; worse, it lacks any sense of urgency. In an amazing display of industry arrogance and lobbying power, the bill provided huge incentives for increased oil and gas production in spite of the industry's record profits as a result of unprecedented energy pricing levels.

In February 2004, Dr. Deffeyes stated, "Because the government has no 'Plan B,' they are essentially gambling that Hubbert and others are wrong, and their inaction will have catastrophic consequences." (M. King Hubbert was the geophysicist who accurately predicted in 1956 that 1970 would be the peak year for U.S. oil production and decline thereafter, and that the same would happen to world production soon after 2000.) Practically every other industrialized nation has attempted to insulate itself from the negative impacts of rising oil prices by employing

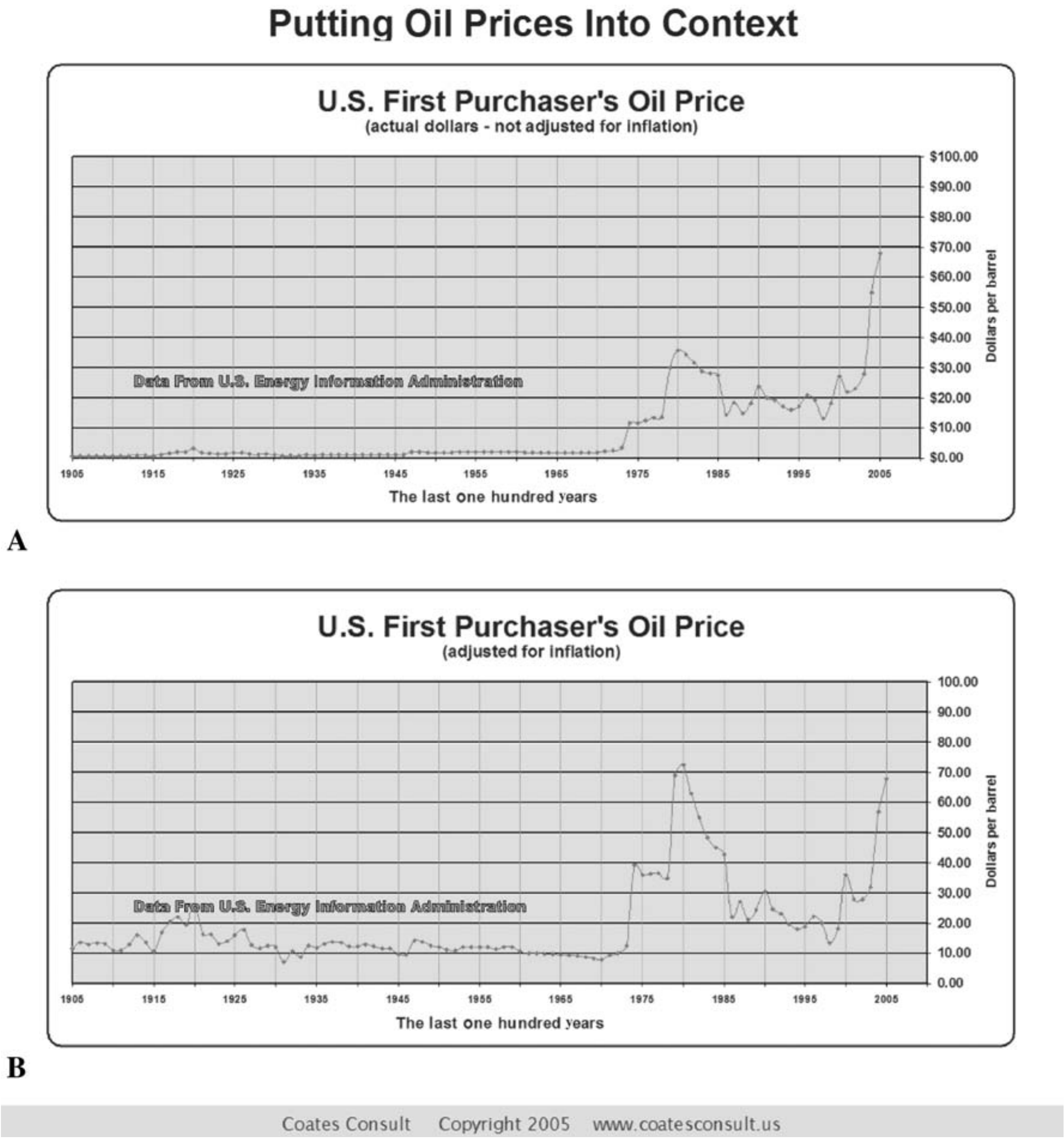


Fig. 7 Putting oil prices into context. (A) These two charts show the difference between real and inflation adjusted oil prices over the last 100 years. There are three outstanding episodes depicted in both charts: a long 65 year period of relative price stablity, followed by two major price spikes. In spite of two world wars and the Korean and Vietnam conflicts, the only significant difference is the dollar valuation in the 1970's. (B) The oil price spike in 1973, due to the first Arab oil the first Arab oil embargo, is obvious. What is not obvious, is that oil production in the United States peaked in 1970, making the country increasingly more dependent on foreign supplies and more vulnerable to volatile world oil markets. The oil embargos of the 1970's were a time of artificial, politically induced shortages. What is occurring in the early 21st century is the beginning of a period of soaring world demand in the face of a constricted oil supply infrastructure. This is a real long term problem and is not artificial. In spite of some claims to the contrary, strong conservation measures are needed.

meaningful energy conservation measures and by extensively deploying electric-powered intercity and intracity transit systems that are not oil reliant.

In light of evolving world events such as China's and India's dramatic economic growth, the world's burgeoning population, the present capacity limits of existing world oil production, the growing concern over the environmental impacts of burning hydrocarbons, and the ominous prospect of terrorist attacks[18] or natural disasters disrupting the flow of oil or natural gas, America needs a strong energy-efficiency program that proactively curtails oil demand and simultaneously supports the alternative development of an extensive electric-powered ground transportation network.

If America's leaders continue to ignore these problems, the country will be left in the weak position of constantly reacting to an increasingly volatile world energy market. The days immediately after the September 11, 2001, attacks exposed the weaknesses in the American transportation system when no practical rail transportation was available for many stranded travelers, and what rail systems were available were overwhelmed until commercial flights resumed several days later.

In spite of this collision course between energy and transportation, America persists in converting vast tracts of productive farmland into distant housing developments because farmland is cheaper than the land nearer to city centers. However, these housing "cost savings" are shifted to the energy and transportation sectors. This "distant affordable housing" and the lack of planned development that encourages energy transportation efficiency add ever more cars to highways that are already filled to capacity and increases the per capita miles driven in America, now double the amount of just 30 years ago.[1] It is this decades-long, automobile-dependent land development policy, combined with the SUV exemption from Corporate Average Fuel Economy (CAFE),[19] that is driving America's astronomical and ever-rising oil use.

Regardless of the many problems facing the nation, the first step is to understand the issues. One single thread running through all of America's more pressing national challenges seems to be a lack of a concerted vision for moving America forward on the eve of significant world change. This change is being spurred by exponential growth in world population and the concomitant growth of energy use in emerging economies. As the world begins to grapple with the impacts of these changes—not only on energy use but also on the quality of the environment—societies will be forced to transition from the previous oil-powered century into a new electric-powered century (Fig 8).[20]

How America generates and transmits its electricity is the pre-eminent challenge of the 21st century. Few Americans think about where their electricity comes from and would likely be astounded to learn that the majority of America's electricity, 51.8%, comes from burning coal.[21] To put things into a human-energy perspective, if a person were to pedal a stationary bicycle attached to a generator that powered a light bulb, the average, reasonably fit adult would be able to light a 100-watt light bulb for about a half hour before exhausting himself or herself. On average, burning 1 lb of coal is enough energy to light ten 100-watt light bulbs for 1 h. An average American's electricity usage consumes 20 lb of coal per day (incidentally, there is about 38% more energy in 1 lb of gasoline, or approximately 1 pint), and an average city with about 1 million people consumes more than 20,000 ton of coal daily—equal to the capacity of 200 railroad coal cars.[22] While tremendous progress has been made to keep emissions low in new coal-fired power plants, coal is still a dirtier hydrocarbon than oil or natural gas. Furthermore, it is also a labor-reliant fuel source with a history of work stoppages that have disrupted the flow of energy to society. It is also reliant on a rail transportation infrastructure that is running at near capacity.

According to a recent Massachusetts Institute of Technology team study,[23] it is likely that nuclear power

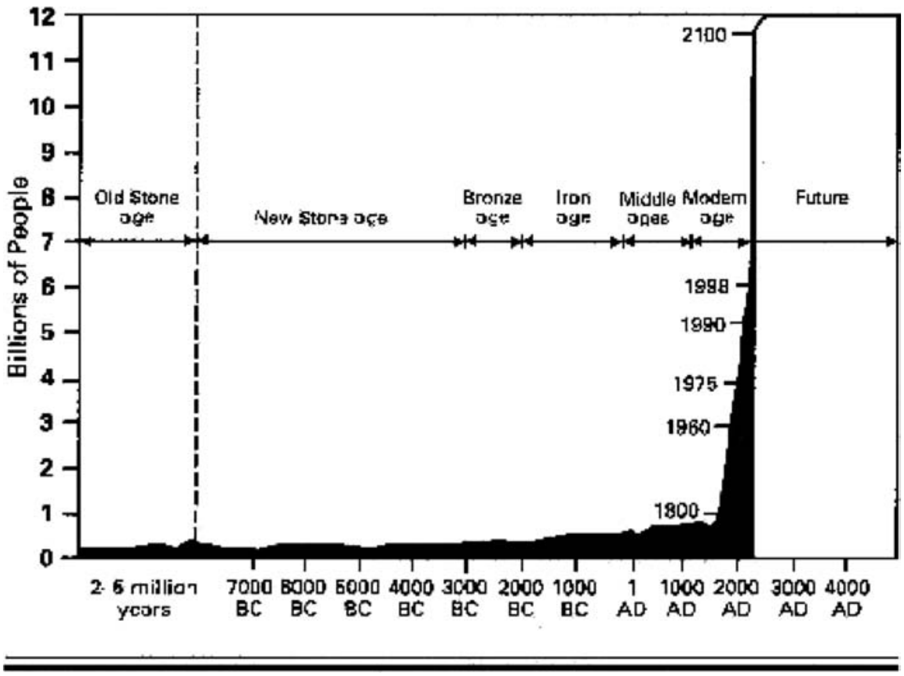


Fig. 8 World population growth history. (World population was estimated at 6.3 billion in 2004 and it is still climbing).
Source: From Population Reference Bureau (see Ref. 26).

will be the only workable solution to meet our prodigious short-term energy needs while preventing greenhouse-gas emissions before cleaner "alternative" or "renewable" energy sources can be developed in sufficient quantities to meet our energy needs. Yet regardless of the source of power used to generate future electricity, there are serious infrastructure problems facing America's method of delivering that electric power. The same problem the oil industry faces with capacity limitations and bottlenecks also plagues the present electricity transmission grid. Transmission in America has been in dire need of massive investment for more than 25 years.

In its present state, the transmission power grid would not be able to accept the prodigious energy demands of converting large segments of America's transportation system to electric power, even if sufficient electricity generation capacity were available. Estimates are that generation capacity would need to triple or even quadruple to convert America's present transportation energy use to electricity. In other words, the grid must be expanded.[24] New transmission technologies such as superconductors and composite core transmission lines will no doubt play a major upgrading role for increasing the existing grid's capacity.

However, regardless of the source of electricity generation, reducing energy usage is the fastest, easiest, and most workable solution. This would necessarily require a renewed focus on better utilization of already available energy. Of course, the objective and the trick are to do this without sacrificing anyone's standard of living or requiring Americans to suffer physical deprivation or any real loss of freedom. Implementing a strategy that demands greater energy efficiency from all sectors of the economy, especially in regard to how and where real estate is developed, will solve many of the country's energy and transportation problems and would likely lead to a higher quality of life for more Americans.

America could implement stricter energy-efficiency requirements and create new employment opportunities by building homes, workplaces, and entertainment centers near interconnected transit hubs (known as Transit-Oriented Development, or TOD). Improved methods for insulating buildings, covering large building surfaces with photovoltaic cells to capture solar energy, and increasing the use of geothermal energy for space heating and cooling rather than burning hydrocarbons makes perfect sense.

For improved transportation efficiency, America can encourage the development and convenient use of reliable, electric-powered, quiet, efficient, and low-maintenance transit for short- and medium-distance travel, such as completely automated maglev systems. For example, the new 267-mph maglev airport connector in Shanghai, China, has been in commercial operation for over a year and is averaging an extraordinary 99.92% on-time-to-the-second schedule for arrivals and departures.[25] The low-speed (60 mph) maglev deployed in Nagoya, Japan, in March 2005 is achieving an astounding 99.97% level of reliability. With extensive electric-powered transit networks in place all along the grid, use of electric-powered cars for short local "last mile" trips would be more convenient and feasible. While rural state residents would not benefit directly from transit development for America's urban areas, their political support for urban transit could reduce overall national oil demand and thus lower the pressure on rising gasoline and oil prices, which are now hurting the agricultural sector. The list of constructive things to be done is almost endless, but by doing nothing new or by simply maintaining the status quo, America is only inviting ever-higher energy prices to force the issue, albeit painfully.

Rather than stubbornly chasing the single-minded obsession with finding cheap oil and internal combustion transportation solutions, it is critical for our federal, state, and local governments to reject their Kmart mentality of pursuing the cheapest infrastructure options and instead to focus on providing the highest-value solutions for present and future generations. And only by adopting a coherent and unified national policy that fully addresses the mutual impacts of energy, transportation, and human settlement patterns will America be able to orchestrate an effective strategy for proactively meeting its future civilian and military energy demands. However, such tectonic shifts in policy will require informed, engaged, and gutsy political leadership.

Access to cheap oil in the 20th century was the key element in America's economic development. Cheap oil was also instrumental in enabling America's successful projection of military might and preserving national security. By aggressively pursuing energy conservation simultaneously with the development of the renewable energy industry, by pursuing land development that is constructed more on the human-powered transportation scale (walking, cycling, and transit), and by converting our oil-reliant transportation system to electricity, America can regain its energy independence, achieve greater national security, create a more stable economy, and make a high quality of life for average citizens more easily attainable and sustainable.

America has the technology and financial capacity to physically transform its infrastructure systems into being more energy efficient. The country has a choice: change its energy attitudes and regain control over the quality of people's lives, or simply remain energy victims, constantly reacting to rising energy prices in a hostile world. It really boils down to choosing and electing politicians who will re-educate themselves on the energy issues, show the courage to speak the truth, and then demonstrate the political will to lead the country in the right direction.

REFERENCES

1. Schrank, D.; Lomax, T. *The 2004 Urban Mobility Report*, Texas Transportation Institute; 2004.
2. U.S. Department of Energy. Energy Information Agency. Transportation Energy Data Book, 23rd Ed., 2003.
3. The Transportation Equity Act for the 21st Century was enacted June 9, 1998, as Public Law 105–178. TEA-21 authorizes federal surface transportation programs for highways, highway safety, and transit for the six-year period 1998–2003.
4. Oberstar, J.L.; *The Transportation Equity Act: Legacy for Users*. Presented at: U.S. House of Representatives: *November 19*; washington, DC, 2003; Vol. 19.
5. U.S. Department of Transportation. Federal Highway Administration. Available at: www.fhwa.dot.gov/hep10/nhs (accessed on February 2005).
6. Mouawad, J. Oil tanker shortages drive costs higher. International Herald Tribune **2004**, *October 20*.
7. Lederer, E.M. *China Loses Faith in Dollar*; Associated Press: Rochester, NY, 2005.
8. U.S. Department of Energy. Energy Information Agency. International World Outlook Report. May 2003.
9. Anderson, H. *US: World Energy Demand to Grow 60 Percent*; United Press International: 2002.
10. Goodman, P.S. Booming china devouring raw materials: producers and suppliers struggle to feed a voracious appetite. Washington Post **2004**, *May 21*, A01.
11. Simmons, M.R. *The World's Giant Oil Fields*; Hubbert Center Newsletter 2002/1, M. King Hubbert Center for Petroleum Studies, Petroleum Engineering Department, Colorado School of Mines: Golden, CO, January 2002.
12. Gerth, J. Forecast of rising oil demand challenges tired saudi fields. New York Times **2004**, *February 24*.
13. Simmons, M. *The Saudi Arabian Oil Miracle*, Matthew Simmons slide presentation, Center for Strategic and International Studies (CSIS): Washington, DC, February 24, 2004.
14. *In Search of Reasonable Certainty: Oil and Gas Reserves Disclosures*, Cambridge Energy Research Associates (CERA): Cambridge, MA, February 2005.
15. Deffeyes, K.S. *Hubbert's Peak: The Impending World Oil Shortage*; Princeton University Press: Princeton, NJ, 2001; p. 22.
16. U.S. Department of Energy. Energy Information Agency. International Petroleum Monthly. Available at: www.eia.doe.gov/emeu/ipsr/t24.xls (accessed on February 2005).
17. OPEC Bulletin, Vol. XXXVI, No. 2. ISSN 0474-6279, February 2005, 70, The Organization of the Petroleum Exporting Countries (OPEC), Vienna, Austria.
18. Baer, R. *Sleeping with the Devil*. Crown Publishers: New York, 2003.
19. U.S. Department of Transportation. National Highway Traffic Safety Administration. Reforming the Automobile Fuel Economy Standards Program. 49 CFR Part 533 [Docket No. 2003-16128], RIN 2127-AJ17. Available at: www.nhtsa.dot.gov/cars/rules/CAFE/rulemaking/ANPRM_Dec22-2003.html (accessed on February 2005).
20. EPRI *Electricity Technology Roadmap: Meeting the Critical Challenges of the 21st Century*; EPRI, 2003.
21. U.S. Department of Energy. Energy Information Administration. *International Energy Outlook*, 2004.
22. U.S. Department of Energy. Energy Information Administration. Coal Data Report 2003.
23. *The Future of Nuclear Power*, An Interdisciplinary MIT Study, Massachusetts Institute of Technology: Cambridge, MA, 2003; ISBN 0-615-12420-8.
24. Grant, P. EPRI, 2004 telephone conversation, and Harding J, FRA, 2004 (in-person interview).
25. Transrapid International-USA, Inc. Transrapid U.S. Reporter Newsletter **2005**.
26. Mofall, J.A., Jr. In *Population: A Lively Introduction, Population Bulletin*, *Vol. 46*; Population Reference Bureau: Washington, DC, 1991; 1–43.

Energy: Global and Historical Background

Milivoje M. Kostic
Department of Mechanical Engineering, Northern Illinois University, DeKalb, Illinois, U.S.A.

Abstract
The global and historical overview of energy use is presented with emphasis on energy diversity but also universality. Starting from ancient civilization a chronology of selected energy-related events is presented. It starts from the prehistoric age, when humans relied on their muscular power to survive; then they learned how to control and use fire, and to domesticate and use animal power, and slowly evolved from hunters and food-gatherers to developers of early agriculture. The use of water and wind power (waterwheels and windmills) expanded human activities and mobility. Further developments included smelting copper and iron ores; using charcoal; and developing different tools, gunpowder, and sailing ships. The use of coal in the mid-1700s and the development of steam engines set off fast growth of cities, population, and further inventions, including internal-combustion engines and the discovery and use of oil, natural gas, and electricity. This accelerated growth period, known as the Industrial Revolution, matured by the end of the 19th century with significant use of fossil fuels and further electrification, and resulted in almost-exponential growth of population and energy use. After the development of nuclear energy and the realization that the abundance of inexpensive fossil fuels will come to an end, along with concern about global environmental pollution, a modern era, with computerization and global Information Revolution, has been taking place. After all developments, life may be happier in the post-fossil fuel era, which represents only a bleep on the human-history radar screen.

INTRODUCTION AND GLOBAL OVERVIEW: ENERGY DIVERSITY AND UNIVERSALITY

The historical overview of energy intrinsically includes geological and societal (human) chronological developments. Energy is more than universal currency. The world view from inside to outside is possible, figuratively and literally, only through the energy prism. From shining stars to rotating planets; to global water, atmospheric, and life cycles; to the evolution, industrialization, and modernization of civilization, energy is the cause and measure of all that there has been, is, and will be.

Each and every material system in nature possesses energy. The structure of any matter and field is energetic, meaning active—i.e., photon waves are traveling in space; electrons are orbiting an atom nucleus or flowing through a conductor; and atoms and molecules are in constant interactions, vibrations, or random thermal motions. Energy is a fundamental property of material systems and refers to the system's potential to influence changes to another system by imparting work (forced directional displacement) or heat (forced chaotic displacement/motion of a system microstructure). Energy exists in many forms: electromagnetic, electrical, magnetic, nuclear, chemical, thermal, and mechanical. Electromechanical energy may be kinetic or potential, whereas thermal energy represents overall chaotic motion energy of molecules and related microstructures. Energy is the cause of all processes across all space and time scales, including global and historical changes. Actually, energy is "the building block" and fundamental property of matter and space; thus, it is a fundamental property of existence, as elaborated in the "Physics of Energy" article in this encyclopedia and elsewhere.[1,2] Energy is both the cause and the consequence of formation and transformation within the universe (everything we are capable of observing or comprehending) at the grand scale, down to the smallest subnanostructures within an atom nucleus and electromagnetic radiation.

Energy warms our planet Earth and keeps it alive. It moves cars and trains, and boats and planes. Energy bakes foods and keeps them frozen for storage. It lights our homes and plays our music. Energy makes our bodies grow and live and allows our minds to think. Through the centuries people have learned how to harvest and use energy in different forms to do work more easily and live more comfortably. No wonder that energy is often defined as ability to perform work—i.e., as a potential for energy transfer in a specific direction (displacement in force direction), thus achieving a purposeful process, as opposed to dissipative (less-purposeful) energy transfer in the form of heat.

Zooming in space and history from the formation of our planet Earth some 4.5 billion years ago, it is observed that our planet has been changing ever since due to energy

Keywords: Energy; Power; Fire; Fossil fuels; Steam and heat engines; Industrial revolution; Electrification; Nuclear energy; Solar energy; Computerization and information revolution; Global environmental pollution.

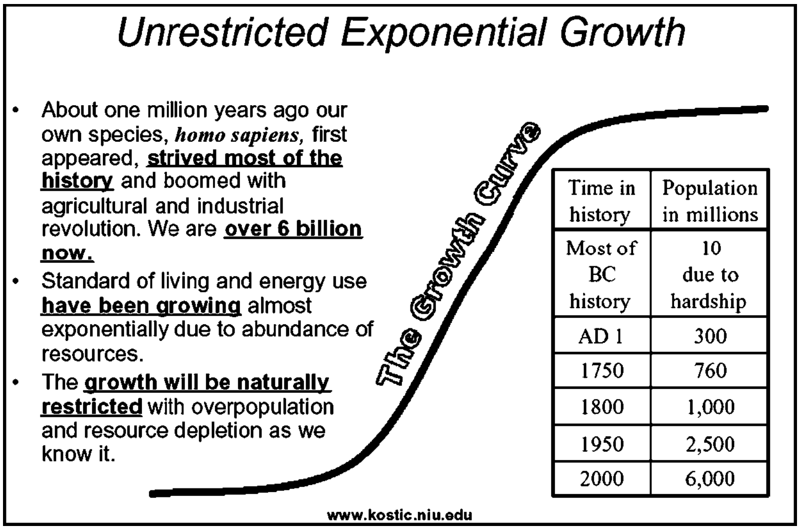


Fig. 1 Population historical growth caused by energy use.

exchanges or "energy flows" in different astrophysical, geological, chemical, biological, and intellectual processes. Hundreds of millions of years ago, life emerged from the oceans and transformed the landscape. Just a few million years ago, the first human species evolved and began its own process of interaction with its environment: the planet Earth. About 1 million years ago our own species, *Homo sapiens*, first appeared, then strived most of the history, and boomed with agricultural and the Industrial Revolution (see Fig. 1).

The current world population is about 6.3 billion. Standards of living and energy use have been growing almost exponentially due to an abundance of resources (see Fig. 2).[3] Today we humans have become sufficiently numerous and technologically active that we may be having a global impact on our planet Earth's environment.[4–6] Growth as we know it, however, will be naturally restricted by overpopulation and resource depletion (see Fig. 1). Two things are certain: in the not-too-distant future (1) the world population and its living-standard expectations will increase substantially, and (2) economical reserves of fossil fuels, particularly oil and natural gas, will decrease substantially. The difficulties that will face every nation and the world in meeting energy needs over the next several decades will be more challenging than what we anticipate now. The traditional solutions and approaches will not solve the global energy problem. New knowledge, new technology, and new living habits and expectations must be developed, both to address the quantity of energy needed to increase the standard of living worldwide and to preserve and enhance the quality of our environment.

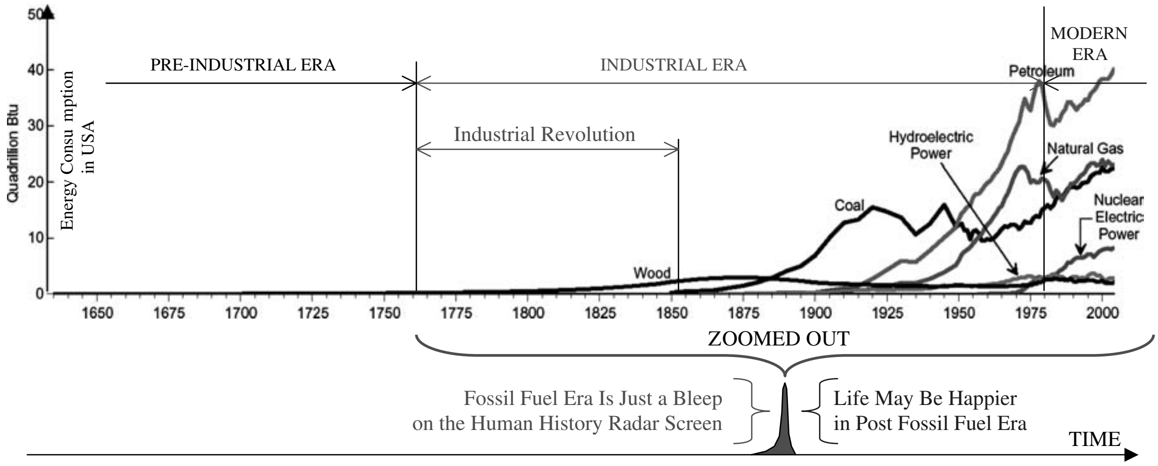


Fig. 2 Energy in history: PreIndustrial era encompasses human evolutionary survival and development of agriculture. Industrial era starts with Industrial Revolution and use of fossil fuels (just a bleep on the human history radar screen). Modern era start with human awareness of fossil fuel depletion and environmental pollution concerns.
Source: From U.S. Department of Energy (see Ref. 3).

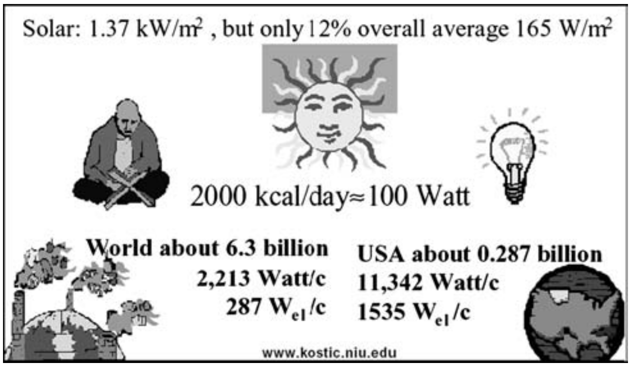


Fig. 3 Energy data: solar energy; dietary metabolic energy; and world and USA energy consumption per capita (per Person).

The human metabolism needed to maintain life is approximately equal to the dietary energy reference value of 2000 kcal/day, which is equivalent to 97 W or 331 Btu/h. Human sustained working power is about 75 W, or one-tenth of 1 hp. Human muscular power bursts may be a hundred times greater than basal metabolic or sustained power. By comparison, the world's population is about 6.3 billion, with total energy consumption about 7550 Btu/h or 2.21 kW per capita (or 11.34 kW per capita for a population of about 0.3 billion in the United States). The total energy rate in kW is often scaled by the typical 33% of thermal-to-electrical conversion efficiency to be compared qualitatively with the electrical energy rate in kW, and vice versa. The corresponding per-capita electricity consumption rate is 0.287 kW and 1.535 kW in the world and the United States, respectively (see Fig. 3).

The total energy coming to the Earth's surface is 99.98% solar, 0.02% geothermal, and 0.002% tidal/gravitational. Currently, about 14 TW (Terawatt, or 2.2 kW/capita [per person]) of the world's energy consumption rate represents only a tiny fraction—0.008%—of the solar energy striking Earth and is about six times smaller than global photosynthesis (all life). Global photosynthesis is only 0.05% of total solar energy, whereas global atmospheric water and wind are about 1% of solar energy. Note that the energy rate or power of 1 TW$=10^{12}$ W$=$29.9 QBtu/year$=$5.89 bbl/year (billion barrels of oil per year) and 1 Quad (QBtu)$=10^{15}$ BTU$=1.055\times10^{18}$ Joule$=$1.055 EJ (Exa-Joule). For energy-unit conversion and fuel, and other energy equivalents, see below and Table 1.

As an ultimate energy source for virtually all natural processes, the solar energy is available for direct "harvest" if needed and is absorbed by vegetation and water surfaces on Earth, thus being the driving force for natural photosynthesis and, in turn, for biosynthesis processes, as well as the natural water cycle and all atmospheric processes. (See the solar-related renewable energy sources in Table 2.) The solar-radiation power density incident to Earth's atmosphere, known as the Solar Constant, is 2 cal/min/cm^2 or about 1.4 kW/m^2—which, after taking into account average day/night time (50%), varying incident angle (50%), and atmospheric/cloud scatter and absorption (53%), reduces to only $0.5\cdot0.5\cdot0.47=11.7\%$ of the Solar Constant, or about 165 W/m^2 at the Earth's surface, as the all-time average (see Fig. 3).

Table 1 Energy units with conversion factors and energy equivalents

Energy units	**J**	**KWh**	**Btu**
1 Joule (J)	1	2.78×10^{-7}	9.49×10^{-4}
1 Kilowatt hour (kWh)	3.6×10^{6}	1	3.412×10^{3}
1 Kilocalorie (kcal$=$Cal$=$1000 cal)	4187	1.19×10^{-3}	3.968
1 British thermal unit (Btu)	1055	2.93×10^{-4}	1
1 Pound-force foot (lb_f ft)	1.36	$3.78\ 10^{-7}$	1.29×10^{-3}
1 Electron volt (eV)	1.6×10^{-19}	4.45×10^{-26}	1.52×10^{-22}
1 Horsepower$\times$second (hp s)	745.7	2.071×10^{-4}	0.707
Energy equivalents	**J**	**KWh**	**Btu**
1 Barrel (42 gal) of crude petroleum	6.12×10^{9}	1700	5.80×10^{6}
1 Ton (2000 lb) of bituminous coal	2.81×10^{10}	7800	2.66×10^{7}
1000 Cubic feet of natural gas	1.09×10^{9}	303	1.035×10^{6}
1 Gallon of gasoline	1.32×10^{8}	36.6	1.25×10^{5}
1 Gram of uranium 235	8.28×10^{10}	2.30×10^{4}	7.84×10^{7}
1 Gram of deuterium	2.38×10^{11}	6.60×10^{4}	2.25×10^{8}
2000 Dietary food calories (2000 kcal)	8.374×10^{6}	2.326	7.937×10^{3}
1 Solar constant$\times$cm$^2\times$sec	8.374	2.326×10^{-6}	7.937×10^{-3}

Source: From Elsevier Inc. (see Ref. 1).

Table 2 Primary energy sources and conversion to work

Primary energy source	Conversion
Non-renewable	
Fossil fuels	
Coal	Combustion (heat and heat-engine H/HE/W[a])
Peat	
Oil/crude petroleum	
Natural gas	
Nuclear	
Uranium	Fission (H/HE/W)
Thorium	
Deuterium	Fusion[b] (H/HE/W)
Renewable[c]	
Geothermal[d]	
Hot steam/water	H/HE/W
Ground soil/rock heat	
Volcanic, etc.[b]	
Ocean-gravitational	
Tidal-ocean wave	Direct to work
Solar-related	
Ocean	
Ocean thermal	H/HE/W
Ocean currents	Direct to work
Ocean wave	
Biomass	
Wood	Combustion (H/HE/W)
Vegetation, etc.[e]	
Direct solar	
Solar-thermal	H/HE/W
Photoelectric	Direct to work
Photochemical	
Electrostatic	
Lightning, etc.[b]	
Wind	
Wind–air streams	
Hydro	
River/accumulation	
Muscular	
Human and animals	

Note: Secondary energy sources (electrical, synthetic fuels, hydrogen, etc.), with energy storage and distribution complete the energy supply domain, which with energy needs, consumption and efficiency complete the global energy realm.

Energy related processes: electromagnetic radiation; photosynthesis cycle in nature; biosyntheses cycle in nature; electrical processes: electro-dynamic, electro-magnetic, electro-chemical; nuclear reactions: fission, fusion, radioactive radiation; chemical reactions: combustion, oxidation, etc.; heat transfer and frictional dissipative processes; thermo-mechanical expansion and compression; natural air streams driven by solar dissipation and gravitational buoyancy (wind); natural water cycle driven by solar dissipation, gravitation and buoyancy (evaporation, precipitations, water streams); natural water streams (rivers and ocean streams); mechanical expansion and compression.

[a]H/HE/W, conversion to Heat and via Heat-Engine to Work.

[b]Not commercialized yet.

[c]All renewables, except tidal and geothermal, are due to solar radiation.

[d]Usually renewable, but may be non-renewable.

[e]Includes many types, as well as waste/garbage.

Source: From Elsevier Inc. (see Ref. 1).

Energy, Work and Heat Units, and Energy Equivalents

Energy is manifested via work and heat transfer, with a corresponding Force × Length dimension for work (N m, kg_f m, and lb_f ft, in SI, metric, and English system of units, respectively); and the caloric units, in kilocalorie (kcal) or British thermal unit (Btu), the last two defined as heat needed to increase a unit mass of water (at specified pressure and temperature) for 1 degree of temperature in their respective units. Therefore, the water-specific heat is 1 kcal/(kg °C) = 1 Btu/(lb °F) by definition, in metric and English system of units, respectively. It was demonstrated by Joule that 4187 N m of work, when dissipated in heat, is equivalent to 1 kcal. In his honor, 1 N m of work is named after him as 1 Joule, or 1 J, the SI energy unit, also equal to electrical work of 1 W s = 1 V A s. The SI unit for power, or work rate, is watt—i.e., 1 J/s = 1 W—and also corresponding units in other system of units, such as Btu/h. Horsepower is defined as 1 hp = 550 lb_f ft/s = 745.7 W. Other common units for energy, work and heat, and energy equivalents for typical fuels and processes are given in Table 1.

Energy Sources

Energy is provided from different sources—i.e., those systems (substances or natural phenomena) that allow for abundant, convenient, efficient, and thus economical conversion of their energy into useful energy forms (for consumption needs). This form usually is thermal for heating, and mechanical and electrical for work, with the latter being also very convenient for transmission and very efficient for conversion into any other useful energy forms. Because energy consumption needs are time and location dependent, energy conversion rate, energy density (per unit mass, volume, area, etc.), transportation (transmission), and storage are important.

There are many sources of energy (see Table 2) that provide for the diverse needs of human activities and society in general. Energy consumption may be classified in four general sectors: (1) residential, for appliances and lighting, space heating, water heating, air-conditioning, etc.; (2) commercial, for lighting, space heating, office equipment, water heating, air-conditioning, ventilation, refrigeration, etc.; (3) industrial, for water and steam boilers, direct-process energy, machine drive, etc.; and (4) transportation, for personal automobiles, light and heavy trucks, air-, water-, pipe-, and rail-transport, etc., see Fig. 4.[7] In all four sectors, in addition to primary energy sources, electrical energy, as a secondary energy source produced from primary energy sources, is used extensively, as presented elsewhere. Conversion efficiencies from different energy sources to useful mechanical or electrical work are given in Table 3.

Chronology of Events

A chronology of selected energy-related events is presented in Table A1 in the appendix.[8–10] It starts from the prehistoric age, when humans relied on their muscular power to survive; then they learned how to control and use fire, and how to domesticate and use animal power, and slowly evolved from hunters and food gatherers to cultivators of crops and developers of early agriculture. The use of water and wind power (waterwheels and windmills) enabled humans to expand their activities and mobility. Further developments included smelting of copper and iron ores, using wood and charcoal, and developing different tools, gunpowder, and sailing ships. The use of coal in the mid-1700s and the development of steam engines set off fast growth of cities, population, and further inventions, including internal-combustion engines and the discovery and use of oil, natural gas, and electricity. This accelerated growth period, known as the Industrial Revolution, matured by the end of the 19th century with significant use of fossil fuels and further developments in electricity, resulting in almost-exponential growth of population and energy use. After the development of nuclear energy and realization that the abundance of inexpensive fossil fuels will come to an end, along with concerns for global pollutions, a modern era, with computerization and global Information Revolution, has been taking place.

Regardless of the depletion of fossil-fuel resources, however, the outlook for future energy needs is encouraging. There are many diverse and abundant energy sources with promising potential, so mankind should be able to enhance its activities, standard of living, and quality of life by diversifying energy sources and by improving energy conversion and utilization efficiencies while increasing safety and reducing environmental pollution.

PREINDUSTRIAL ERA: SURVIVAL AND AGRICULTURAL DEVELOPMENT

In contrast to today's mostly sedentary lifestyle, our ancestors spent most of their existence as hunters and food gatherers, with strong physical and mental challenges to succeed in survival. Those challenges and longtime adaptations ultimately evolved in the complexities of today's societies. It took about 1 million years for our own species, *Homo sapiens*, to survive, literally in hardship, and in most of BC history, the world population was below 10 million. Except for very few early communities living in "favorable" localities, most of our ancestors were surviving on grasslands and forests with population densities comparable to their roaming foragers. Development of traditional agriculture was followed by a rise in population; further cultivation of

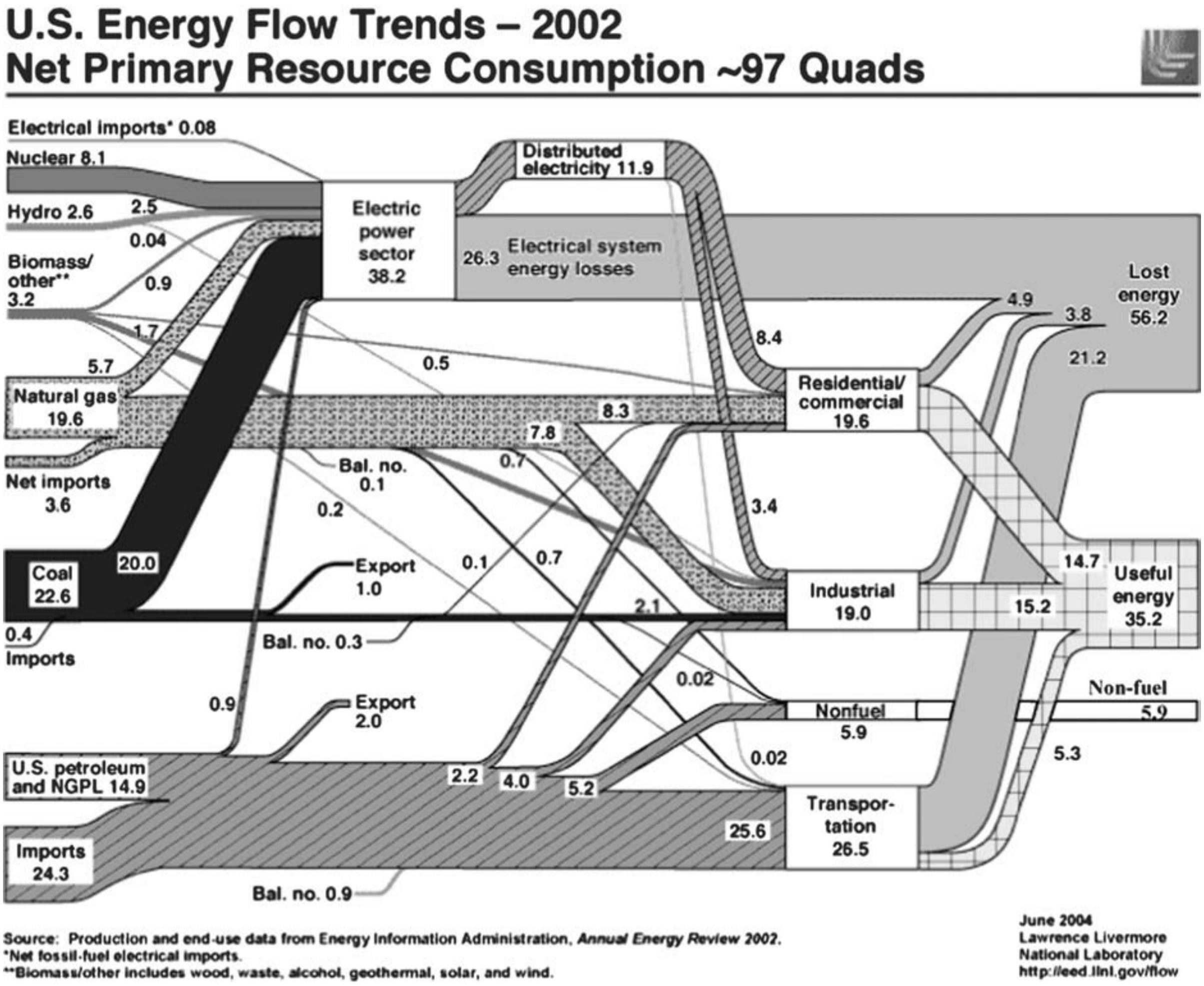


Fig. 4 Energy input–output cross-paths from primary sources to consumption sectors and energy losses. [Note that total energy is close to 100 QBtu, thus numbers are close to %].
Source: From Lawrence Livermore National Laboratory (see Ref. 7).

crops; and domestication of animals, including horses. Many cattle breeds provided draft and power, as well as milk. Virtually all fuel in preindustrial societies came from straw, wood, and charcoal. The latter was critical for smelting and processing, first metals (copper, iron, and steel) and then firing bricks. The power was provided by the muscular labor of people and animals. Even today, in undeveloped rural areas of Asia, Africa, and Latin America, most of the work is provided by human and animal labor.

Smelting metal ores required large quantities of wood and charcoal, as well as skills to sustain high temperatures in metallurgical pits and furnaces. In turn, improved tools and utilities were made of metals, leading to the development of waterwheels and windmills, as well as wheeled carts and sailing ships (see Table A1).[8,9] Increased mobility on land and sea helped in exchange of goods and skills from one area to another, which in turn helped the development of better goods, new materials and tools, and ultimately the rise of population.

INDUSTRIAL ERA: THE FOSSIL FUELS' BLEEP ON THE CIVILIZATION RADAR SCREEN

Development of prime movers using heat from fuels—the heat engines—was a critical historical event, because stored high-density energy in fuels like wood, and particularly coal and oil, could provide energy at any time and in any place. The abundance of fossil fuels (coal, oil, and natural gas) and energy independence from locality and seasonal natural phenomena, such as waterfalls and wind, opened many opportunities for unforeseen development. Invention of the first practical steam engine by Newcomen and Savery in 1712 and improvements by James Watt in 1765 started intensive development and utilization of fossil fuels—still the most dominant energy source, with an 85% share of the total energy use of modern society. The so-called Industrial Revolution was set in motion, with unprecedented developments, including internal-combustion engines; electrification and electrical motors; new devices, materials, and chemicals;

Table 3 Energy conversion efficiencies

Engine/process	Efficiency %
Otto (gasoline) engine	20–30
Diesel engine	30–40
Gas turbine	30–40
Steam turbine	35–45
Nuclear, steam turbine	30–40
Combined gas/steam turbines	40–60+
Fuel cell (hydrogen, etc.)	40–60+
Photovoltaic cell	10–25
Geothermal plant	5–15
Windmill	30–40 (59% limit)
Hydro turbine	80–85
Electro-mechanical motor/ generator	70–98

Note: Thermal-to-mechanical work conversion is limited by stochiometric combustion temperature and the Carnot cycle efficiency. Fuel cell efficiency is limited by Gibbs free energy values for process reactants and products, and may be close to 100%. Due to material property limitations and process irreversibilities (dissipation of energy), practical efficiencies are much lower and there is room for substantial improvements. For example, existing hybrid cars have 80% improved efficiency (and mileage) over the same classical cars, from 25 to 45%, by using electro/mechanical engines/storage hybrid systems.

and other inventions (see Table A1). The birth and intense development of the new energy science, thermodynamics, was taking place, along with the discovery of the fundamental laws of nature and many other discoveries in chemistry and physics. One invention was fueling another invention, and so on. The use of new heat engines and the need for more fuels were propelling discovery of many coal mines and oilfields. In return, available energy sources were enabling an intense rise in human activities, skills, and knowledge, as well as the growth of civilization, reaching 1 billion people by the end of the 18th century (see Fig. 1). The Industrial Revolution matured and continued with the Industrial Era and ultimately evolved into the modern era of societal development.

MODERN ERA: SOPHISTICATION, CONSERVATION, AND DIVERSIFICATION

The Modern Era in societal development represents a continuation of the Industrial Era, with development of new technologies (including nuclear energy, space exploration, computerization, and information technologies) as well as the realization that the abundance of inexpensive fossil fuels will come to an end, along with concern about global environmental pollution.

The primary energy sources for the world in 2003 and the United States in 2004 are presented in Table 4, and the primary sources for the production of electricity are presented in Table 5.[3,4] In addition, the U.S. energy supply by consumption sector, including electricity production, is given in Table 6. The world and U.S. populations, energy production, and consumption also are summarized in Table 4. Total energy production—including losses, import, and export—is available as energy supply for consumption and storage. Also, most of the world's electricity (about 65%, and about 71% in the United States) is produced from fossil fuels, with overall conversion efficiency of only about 33%. Conversion

Table 4 World and U.S. total energy supply by source (in QBtu)

Source	World, 2003		U.S., 2004	
Coal	99.69	23.9%	22.528	22.6%
Petroleum	159.17	38.2%	40.130	40.2%
Natural gas	98.7	23.7%	22.991	23.1%
Fossil fuels	357.56	85.7%	85.649	85.9%
Nuclear electric	26.52	6.4%	8.232	8.3%
Hydro-electric	27.18	6.5%	2.725	2.7%
Renewables/others	5.87	1.4%	3.391	3.4%
Total	417.12	100.0%	99.740	100.0%
World and U.S. population and energy comparisons				
Population	6,300	100%	294	4.7%
Energy production	417.12	100%	70.369	16.9%
Energy consumption	417.12	100%	99.740	23.9%

Note: Energy in Quadrillion Btu (1 QBtu $=10^{15}$ Btu) or %, population in Millions.
Source: From U.S. Department of Energy (see Refs. 3 and 4).

Table 5 World and U.S. electric energy supply by source (in Billion kWh)

Source	World, 2003		U.S., 2004	
Coal			1,976.3	50.0%
Petroleum			117.6	3.0%
Natural Gas			714.6	18.1%
Fossil Fuels	10,364.8	65.4%	2,808.5	71.0%
Nuclear Electric	2,523.1	15.9%	788.6	19.9%
Hydro-electric	2,645.8	16.7%	269.6	6.8%
Renewables/Others	241.9	1.5%	89.2	2.3%
Total	15,843.9	100%	3,955.9	100%

Note: Energy in Billion kWh or %; 1 kWh(electric) equivalent to 10580 Btu(thermal) at 33% efficiency, however 1 kWh = 3412 Btu (as unit conversion).
Source: From U.S. Department of Energy (see Refs. 3 and 4).

efficiency is similar in nuclear power plants, which contribute to about 16% of world and about 20% of U.S. electricity production. When the global energy supply is given together with fossil fuels and expressed in British thermal units (Btu), all electrical energy (including hydro and wind) is given in equivalent Btu thermal units, accounting for the conversion efficiency (typically, 33%). When electrical energy is accounted separately, the actual electrical output is given in kilowatt hours (kWh), as shown in Table 5. Due to different forms and conversion efficiencies of primary energy sources, and due to the complexities of energy production and losses, transportation and storage, import, and export, it is virtually impossible to account correctly for all energy paths and forms in the same units; therefore, the total figures (and percentages) usually do not add up exactly (see Fig. 4 and Table 6 for examples).

Fossil fuels account for more than 85% of total world and U.S. energy consumption (see Table 4). Almost 40% of total world and U.S. primary energy is used for electricity production (see Tables 4–6), mainly in thermal and nuclear power plants (more than 80% in the world and more than 90% in the United States), using heat engines undergoing thermomechanical conversion processes with relatively low conversion efficiencies (see Table 3). The overall conversion efficiency from chemical or nuclear fuel energy to thermal energy of combustion gases or steam, to mechanical and electrical energy, is only about 30%–35%.

FUTURE ENERGY OUTLOOK: LIFE MAY BE HAPPIER AFTER FOSSIL FUELS

At present, most of the world's energy consumption is supplied by fossil fuels (about 85%). The proven fossil-fuel reserves are limited, however, and if they continue to be used at the present rates, it is estimated that coal (as used under current conditions) will be depleted in about 250 years; oil, in 60 years; and natural gas, in about 80 years. We have to keep in perspective that "proven reserves" refers to the customary and economical mining and utilization of fuels, but new reserves and more efficient technologies are being discovered, making new fuel reserves economical. At present, a substantial amount of the world's electricity is obtained from nuclear and hydro energy (about 16 and 17%, respectively), and the use of other renewable energy resources is increasing—namely, geothermal, wind, biomass, and solar. In addition, alternative synthetic fuels, including hydrogen, are being developed. It is worth noting that some countries (including Norway, Brazil, New Zealand, Austria, and Switzerland) produce almost all or most of their electricity from hydro energy, and France produces most of its electricity (more than 75%) from nuclear. Reserves of nuclear fuel are orders of magnitude higher than reserves of fossil fuels, and nuclear fuel does not contribute to CO_2 and greenhouse pollution.

Table 6 U.S. energy consumption by sector in 2004 (in QBtu)

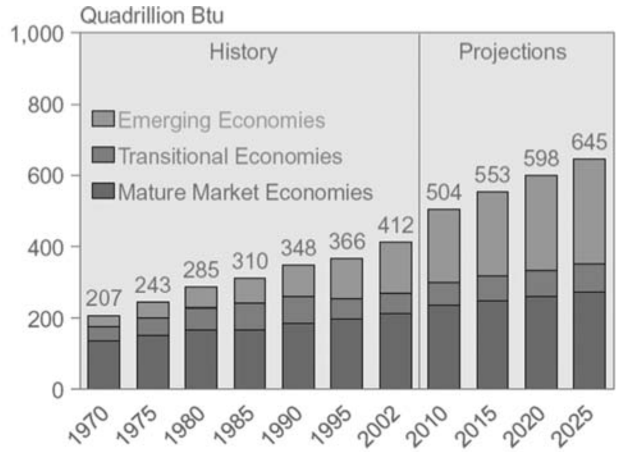
Sector	Primary		Electric		Total	
Residential	7,022	7.1%	14,154	36.43%	21,176	21.2%
Commercial	4,072	4.3%	13,443	34.60	17,515	17.6%
Industrial	22,076	22.3%	11,171	28.75%	33,247	33.3%
Transportation	27,709	27.6%	84	0.22%	27,793	27.9%
Electric	38,850	38.7%				
Total	99,729	100.0%	38,852	100%	99,740	100%

Source: From U.S. Department of Energy (see Ref. 3).

Furthermore, advances in energy conversion and utilization technologies, and increases in efficiency, including computerized control and management, contribute to energy conservation, an increase in safety, and a reduction of related environmental pollution. Actually, per-capita energy use in the United States and other developed countries has been reduced in recent years. The increase of the world's population, however, and the development of many underdeveloped and very populated countries (China, India and others) will influence continuous increase of the world's energy consumption.

Fig. 5 gives one of the most recent projections of the world's energy consumption, by region, until 2025.[5] The Mature Market Economies region (15% of the 2005 world population) represents North America, Western Europe, and Mature Market Asia (Japan, Australia, and New Zealand). The Transitional Economies region (6% of the 2005 world population) represents Eastern Europe (EE) and the former Soviet Union (FSU). The rest is the Emerging Economies region (78% of the 2005 world population), consisting of emerging Asia (53% of the 2005 world population), the Middle East (4% of the 2005 world population), Africa (14% of the 2005 world population), and Central and South America (7% of the 2005 world population).

World energy consumption is projected to increase by 57% from 2002 to 2025. Much of the growth in worldwide energy use is expected in the Emerging Economies countries. The increase is projected on average to be 2.0% per year over the 23-year forecast (from 2002 to 2025)—somewhat lower than the 2.2% average annual growth rate from 1970 to 2002. Worldwide, total energy use is projected to grow from 412 QBtu (quadrillion British thermal units) in 2002 to 553 QBtu in 2015 and 645 QBtu in 2025 (see Fig. 5).[5] Emerging Economies will account for much of the projected growth in energy consumption over the next two decades, with energy use in the group more than doubling by 2025 due to strong projected economic growth in the region. The world population is expected to grow on average by 1% per year (0.4, −0.2, and 1.2% in the Mature, Transitional, and Emerging regions, respectively) and to reach 7.85 billion by 2025. The gross domestic product (GDP) is expected to increase by 3.9% per year on average: 5.1% per year in the Emerging Economies countries, compared with 2.5% per year in the Mature Market Economies countries and 4.4% per year in the Transitional Economies countries of Eastern Europe and the former Soviet Union (EE/FSU). The long-term projections are more uncertain because future development may turn in many different, even unexpected, directions.



Sources: **History:** Energy Information Administration (EIA), *International Energy Annual 2002*, DOE/EIA-0219(2002) (Washington, DC, March 2004), web site www.eia.doe.gov/iea/. **Projections:** EIA, System for the Analysis of Global Energy Markets (2005).

Fig. 5 World energy history and projection consumption by region.
Source: From U.S. Department of Energy (see Ref. 5).

As already stated, two things are certain: In the not-too-distant future (1) the world population and its living-standard expectations will increase substantially, and (2) economical reserves of fossil fuels, particularly oil and natural gas, will decrease substantially. The difficulties that will face every nation and the world in meeting energy needs over the next several decades will be more challenging than what we anticipate now. The traditional solutions and approaches will not solve the global energy problem. New knowledge, new technology, and new living habits and expectations must be developed, both to address the quantity of energy needed to increase the standard of living worldwide and to preserve and enhance the quality of our environment.

A probable scenario, in the wake of a short history of fossil fuels' abundance and use (a bleep on the human-history radar screen), the following energy future is anticipated:

1. Creative adaptation and innovations, with change of societal and human habits and expectations (life could be happier after the fossil-fuels era).
2. Intelligent, high-tech local and global energy management in a wide sense (to reduce waste, improve efficiency, and improve the quality of the environment and life).
3. Unforeseen large (higher order of magnitude) potential for energy conservation and regeneration in industry, transportation, and the commercial and residential sectors.
4. Nuclear energy and re-electrification for most stationary energy needs.
5. Cogeneration and integration of power generation and new industry on a global scale (to close the cycles at sources, thus protecting the environment and increasing efficiency).

6. Renewable biomass and synthetic hydrocarbons for fossil-fuel replacement (mobile energy, transportation, and chemicals).
7. Advanced energy storage (synthetic fuels, advanced batteries, hydrogen, etc.).
8. Redistributed solar-related and other renewable energies (to fill in the gap).

In conclusion, life may be happier after the fossil fuel era, which represents only a bleep on the human-history radar screen. With increased population and technological developments, and sophistication in many areas of complex societies, there will be many unforeseen opportunities to enhance efficiencies of energy production and utilization. Therefore, the outlook for energy needs is encouraging. There are many diverse and abundant energy sources with promising potential, so the mankind should be able to enhance its activities, standard of living, and quality of life by diversifying energy sources and by improving energy conversion and utilization efficiencies, while at the same time increasing safety and reducing environmental pollution.

APPENDIX A

Table A.1 Chronology of selected energy-related events in history

Year	Event in energy history
500,000+ or BC	Middle Pleistocene humans control fire (burning wood). Direct evidence was found outside a cave at Chou k'ou-tien, China, where charcoal was found along with traces of a stone tool
10,000+	Paleo-Indians used hot springs in North America for cooking, and for refuge and respite
6,000+	The earliest known use of ships comes from Egyptian rock drawing dating from 6,000 BC
4,500+	Egyptians mine copper ores and smelt them
4,000+	Horses are ridden in what is now the Ukraine
3,500+	Wheeled vehicles are used in Mesopotamia as seen in a pictograph found in Uruk
1,000+	Coal from the Fu-shun mine in northeastern China may have been used to smelt copper
900+	The use of natural gas was mentioned in writings from China
480+	The Persians used incendiary arrows wrapped in oil-soaked fibers at the siege of Athens
400+	Greek philosopher Democritus theorized that matter consists of tiny particles called atomos, that could not be divided
250+	Archimedes invents a number of items including the Archimedian screw—a helix-shaped screw in a tube for lifting water. He is also credited with having discovered the principles of the lever
211+	It was in China that the first known well was drilled for natural gas to reported depths of 150 m (500 ft)
100+ BC or +	In Illyria (ex-Yugoslavia and Albania), and probably in western Anatolia (Turkey), water-powered mills are used for grinding grain
100 AD	Hero of Alexandria invents the first steam engine called the aeolipile. It consisted of a spherical vessel fitted with two jets pointing in opposite directions. Hero also invented a wind device
300	Chinese learn to use coal instead of wood as fuel in making cast iron
300	Water mills appear in Roman Empire
300	First known references to a perpetual motion machine appears in a Sanskrit manuscript. It describes a wheel with sealed cavities in which mercury would flow in such a fashion that one half of the wheel would always be heavier, providing continuous spinning
600	The earliest known references to wind-driven grain mills, found in Arabic writings
1100	As a result of the Arab invasion of Spain, the industrial art of distillation of petroleum products into illuminants became available in western Europe by the 12th century
1200	The first documented proof that coal was mined in Europe, provided by the monk Reinier of Liège
1200	Alcohol is first distilled in Europe from grains
1221	Chinese use bombs and other uses of gunpowder, leading eventually to development of rockets
1500	Leonardo da Vinci invents many devices including a glider, parachute and a helicopter type of device
1570	William Gilbert studies magnetism and the corresponding attraction of rubbed amber and various rubbed jewels
1603	Hugh Platt discovers coke, a charcoal-like substance produced by heating coal
1609	The first attempt is made to harness ocean energy in the Bay of Fundy

(Continued)

Table A.1 (*Continued*)

1612	A primitive thermometer invented by Galileo
1650	Otto van Guericke develops a way to charge a ball of sulfur with static electricity. He also observed light produced by electricity. The term electricity is coined to describe the force that is found when amber is rubbed with silk (static electricity)
1659	Natural gas is discovered in England; the first time in Europe
1660	Robert Boyle presented a law (Boyle's law), which states that pressure varies inversely with volume at constant temperature, paving the way for the ideal gas law
1670	Christian Huygens builds a motor driven by explosions of gunpowder
1680	Sir Isaac Newton proposes that a jet of steam could be used (like a rocket) to power a carriage, an idea now considered to be a precursor to development of the jet engine
1680	The match is first discovered by Robert Boyle who rubbed phosphorus and sulfur together
1687	In famous Principia, one of the most important and influential works on physics of all times, Isaac Newton presents the universal law of gravitation and the three fundamental laws of motion
1690	The recycled paper manufacturing process is introduced. The Rittenhouse Mill near Philadelphia made paper from fiber derived from recycled cotton and linen rags
1698	Dennis Papin describes an apparatus (called the Papin Cylinder in which the condensation of steam in a cylinder creates a vacuum. He later in 1698 develops the first piston that is moved by the pressure of steam rather than atmospheric pressure
1709	Sir Issac Newton builds an electric generator consisting of a rotating glass sphere
1712	Thomas Newcomen in collaboration with Thomas Savery build the first practical steam engine to use a piston and cylinder
1738	Daniel Bernoulli published the conservation of live forces in hydraulics (now known as Bernoulli equation) and the kinetic molecular theory of gasses
1747	Benjamin Franklin describes in a letter his discovery that a pointed conductor (Franklin's lightning rod) can draw electric charge from a charged body
1755	Leonhard Euler equations for the motion of inviscid incompressible fluid and contributions in mathematics, optics, mechanics, electricity, and magnetism
1765–1776	The steam engine is perfected by James Watt with condenser that is separated from the cylinder
1766	The element hydrogen (symbol is H) is discovered by Henry Cavendish (1731–1810), an English physicist and chemist
1774	Joseph Priestley and Karl Scheele independently discover the element oxygen (symbol is O)
1775	Allesandro Volta describes his electrofore perpetuo (electrophorus), a device for producing and storing a charge of static electricity, replacing the Leiden jar and eventually leads to modern condensers
1777	The first buildings in France since Roman times are heated by warm water central heating systems
1779	The first versions of the bicycle appear in Paris
1783	Montgolfier brothers of France create the first hot air balloons using flame fires
1787	First steamboat in America is demonstrated on the Delware River in Philadelphia, Pennsilvania
1789	The element Uranium (symbol is U) is discovered by Martin Klaproth, a German chemist
1789	Antoine Lavoisier presented a unified view of new theories of chemistry, contained a clear statement of the law of mass conservation, and denied the existence of phlogiston. He lists elements, or substances that could not be broken down further, which included oxygen, nitrogen, hydrogen, phosphorus, mercury, zinc, and sulfur. His list, also included light, and caloric (heat fluid), which he believed to be material substances. He underscored the observational basis as opposed to reasoning
1799	Alessandro Volta creates the first electric battery called the Voltaic pile
1804	Richard Trevithick develops a steam locomotive that runs on iron rails
1816	Robert Stirling invents a power cycle with heated air that operates without a high-pressure boiler
1821	Michael Faraday reports his discovery of electromagnetic rotation. He creates the first electrical "motors," although his rotating needle is not a real motor because it cannot power anything
1821	Johann Seebeck observes that two different metals joined at two different places kept at two different temperatures will produce an electric current. This is called thermoelectricity and the Seebeck effect will later be used in the development of the semiconductor

(Continued)

Table A.1 (*Continued*)

1823	Methyl alcohol was first discovered by condensing gases from burning wood into a liquid. It is used as a solvent and a chemical building block to make consumer products as plastics, plywood and paint
1824	In "On the motive power of fire," Sadi Carnot shows that work is done as heat passes from a high to a low temperature. He defines work and hints at the second law of thermodynamics
1827	George Ohm writes "The galvanic circuit investigated mathematically," which contains the first statement of the Ohm's law, that the electrical current is equal to the ratio of the voltage to the resistance
1830	The first locomotive in the U.S. to carry 26 passengers 13 miles over the tracks of the Baltimore and Ohio Railroad. George Stephenson's steam locomotive was chosen over three competitors to open the Liverpool to Manchester railway in England. This is considered the start of the railroad boom
1831	Michael Faraday independently discovers that electricity can be induced by changes in an electromagnetic field. Though Henry found it earlier, but didn't publish it, so Faraday is credited with the discovery
1832	Joseph Henry discovers self-induction, or inductance, the second current in a coil through which one current is passing that is induced by the first current
1838	The steamship Sirius is the first ship to cross the Atlantic on steam power alone, taking 18 days and very nearly running out of coal before reaching New York
1839	The first work done with photovoltaics was performed by Edmond Becquerel
1839	William Grove develops the first fuel cell, a device that produces electrical energy by combining hydrogen and oxygen
1841	Frederick de Moleyns obtains the first patent for an incandescent lamp, an evacuated glass containing powdered charcoal that bridges a gap between two platinum filaments
1842	Julius Robert Mayer is the first to state the law of conservation of energy (known as the first law of Thermodynamics), noting that heat and mechanical energy are two aspects of the same thing
1943	Joule experimentally measure the "heat equivalent of work," also known as "mechanical equivalent of heat," thus discrediting the caloric theory
1948	William Thomson (i.e. Lord Kelvin) proposed an absolute scale of temperature. It was based on theory proposed by Sadi Carnot and later developed by Clapeyron
1850	Rudolf Clausius and later Rankine formulate the first law of energy conservation, which starts new science "Thermodynamics." The first statement of the second law of thermodynamics was enhanced by Clausius in 1865 as "entropy always increases in a closed system." In other words, other energies in an isolated (i.e., closed) system will change toward heat and disorder
1851	William Thomson (i.e. Lord Kelvin) describes heat pump, a device where heat is absorbed by the expansion of a working gas and given off at a higher temperature in a condenser
1853	William Rankine, proposed a thermodynamic theory with Kelvin based on the primacy of the energy concept. He stated the Law of Conservation of Energy as "all different kinds of physical energy in the universe are mutually convertible." He also invented an absolute temperature based on the interval of one degree Fahrenheit termed the Rankine temperature scale
1853	Kerosene is extracted from petroleum for the first time
1855	Henry Bessemer introduces the Bessemer process for producing inexpensive steel in a blast furnace
1859	Gaston Plante in Paris invents the first lead-acid storage battery, which produces electricity from a chemical reaction and can be recharged again and again
1865	Natural gas first found near Stockton, California when workmen drilling for water found natural gas at 1,800 feet. It supplied gas for lighting the courthouse as well as warm water for nearby swimming baths
1867	Nicholaus Otto developed an internal combustion engine that is an improved version of Lenoir's engine
1870	Heinrich Hertz used Edmond Becquerel's discoveries that certain materials, such as selenium, produced small amounts of electric current when exposed to light. Not long after that, selenium photovoltaic cells were converting light to electricity at 1–2% efficiency
1875	Alfred Nobel accidentally discovered that nitroglycerine retain its explosive properties when absorbed by diatomaceous earth, calling it "dynamite." He left his estate to establish famous "Nobel Prizes."
1882	Joseph Fourier developed a mathematical theory of heat in terms of differential equations
1877	Nikolaus Otto develops the four-cycle internal combustion engine, similar to what we use today
1879	Thomas Alva Edison invented the first electric incandescent lamp of practical value

(Continued)

Table A.1 (*Continued*)

1880	Pierre Curie discovers the piezoelectric effect that certain substances produce electric current under pressure
1882	The first electric central station to supply light and power was the Edison Electric Illuminating Company of New York City. It had one generator which produced power for 800 electric light bulbs. Within 14 months, the service had 508 subscribers and 12,732 bulbs
1884	Nikola Tesla invents the electric alternator, an electric generator that produces alternative current
1885	James Prescott Joule builds an internal combustion engine that is the precursor to the diesel engine
1886	Henry Ford builds his first automobile in Michigan
1888	Heinrich Hertz detects and produces radio waves for the first time. Radio waves are called Hertzian waves until renamed by Guglielmo Marconi, who calls them radiotelegraphy waves
1890	Clement Ader's Eole is the first full-size aircraft to leave the ground under its own power, carrying its inventor as a pilot. The plane crashed on landing, so the invention was not credited to Ader
1891	Nikola Tesla invents the Tesla coil, which produces high voltage at high frequency
1893	Rudolf Diesel describes an internal combustion engines that will be named after him. The ignition of the injected fuel in the cylinder is self-ignited due to high temperature of the compressed air
1895	On the Willamette River at Oregon City, Oregon, the first dam is specifically built to drive a hydroelectric power plant
1897	Joseph Thomson discovers the electron, the particle that makes electric current and the first known particle that is smaller than an atom
1900	The first offshore oil wells are drilled
1901	Guglielmo Marconi transmits first long distance communication using electromagnetic or radio waves
1902	America's first moving-picture theater opens in Los Angeles, charging 10 cents for a one hour show
1902	Maie and Pierre Curie discover the atomic weight of radium
1902	Willis H. Carrier invents the first air conditioner, although his name is first used in 1906 to describe a different device
1905	Albert Einstein published his paper on the photoemissive photoelectric effect, along with a paper on his theory of relativity. The paper describes light for the first time as being both a wave and a particle. He later wins Nobel Prize in 1921
1908	Polish scientist Czochralski developed a way to grow single-crystal silicon, a necessary step for computer industry and solar cells
1919	British airship R-34 is the first airship to cross the Atlantic Ocean
1921	A German scientist, Friedrich Bergius succeeds in liquefying coal into oil in Stuttgart
1927	First successful long-distance television transmission was demonstrated, from President Hoover's office in Washington, DC, to the Bell Laboratories in New York
1929	The first "talking" movie filmed entirely in color was released
1929	Felix Wankel patents a rotary engine, but the engine is not practical until the 1950s
1929	Georges Claude develops the first electrical power plant to use the difference in temperature between the upper and lower layers of ocean
1937	Frank Whittle and A.A. Griffitsh build the first working jet engine in England. Independently in Germany, von Ohain and M. Muller develop a similar engine
1942	The first self-sustaining nuclear chain reaction was demonstrated by Enrico Fermi and his staff at the University of Chicago, making possible the development of the atomic bomb
1945	U.S. explodes the first nuclear weapon at Alamogordo, N.M, a different type of atomic bomb is dropped on the Japanese city of Hiroshima, followed by another nuclear bomb on Nagasaki on August 9, 1945
1946	The ENIAC computer is demonstrated to scientists and industrialists; even though it was used during W.W.II. It multiplies 360 ten-digit numbers and extracts a square root "in a single second."
1946	First Soviet nuclear reactor goes into operation
1947	The first peacetime nuclear reactor in the U.S. starts construction at Brookhaven, NY
1947	Andrei Sakharov and F.C. Frank propose the use of negative muons to produce fusion reactions in a mixture of deuterium and hydrogen. This possibility is rediscovered in 1957 by Luis Alvarez

(Continued)

Table A.1 (*Continued*)

1947	The first airplane to break the speed of sound, the Bell X-1, is flown by Chuck Yeager
1948	The first transistor, invented by Drs. John Bardeen and Walter Houser Brittain, was demonstrated. The essential element of the device was a tiny wafer of germanium, a semi-conductor
1948	Market acceptance of frozen orange concentrate leads to the expansion of the frozen foods industry, with associated increases in packaging
1951	An announcement was made of a battery that converts nuclear energy to electrical energy. Philip Edwin Ohmart of Cincinatti, Ohio, invented the radioactive cell
1952	Westinghouse Electric Corporation builds the first breeder reactor at the U.S. Atomic Energy Commission's laboratories in Arco, Idaho. It produces more plutonium than the uranium it burns, promising an era of cheap nuclear energy
1954	Bell Lab's Chapin, Fuller, Pearson: AT&T, patent "Solar Energy Converting Apparatus," submitted to U.S. patent office
1957	The Soviet Union launches the world's first artificial satellite, Sputnick 1. Another satelite Sputnick 2 carried a live dog Laika into space, but the dog did not return to earth
1958	Integrated circuits produced by Texas Instruments
1959	Francis Bacon builds a fuel cell in which hydrogen and oxygen react in a mixture of potassium hydroxide in water to produce electricity
1961	Yuri Gagarin becomes the first human in space making a single orbit of the planet in 1 h 48 min
1962	Telstar satellite launched; the first commercial telecommunication satellite; project of Bell Telephone Laboratories, proposed in 1955 by John R. Pierce
1969	Humans first steps on the Moon. Neil Armstrong and Michael Collins and Edwin Aldrin on the Apollo 11
1969	Geologists discover oil in Alaska's north slope
1971	Silicone chips are produced by Intel and Texas Instruments
1972	The first electric power using municipal refuse as a boiler fuel was generated by the Union Electric Company's Meramec Plant in St. Louis, Missouri
1973	The polyethylene terephthalate plastic bottle is patented by chemist Nathaniel Wyeth which soon began to replace glass bottles. The recycling of the plastic soon followed
1976	Unmanned US Voyager spacecraft lands on Mars
1977	Trans-Alaska oil pipeline opens
1980s	Personal computers take off
1982	CD players produced by Philips and Sony
1983	Specially built 1-kW, PV-powered car, the Solar Trek, drives across Australia, covering 4000 km in less than 20 days. The maximum speed was 72 kph, with an average speed for the trip of 24 kph
1984	A 20-MW geothermal plant opened at Utah's Roosevelt Hot Springs
1985	Martin Green team, University of New South Wales, Australia, breaks the 20-percent efficiency barrier for silicon solar cells under "1-sun" conditions
1986	Chernobyl nuclear reactor number 4 near Kiev in ex-Soviet Union explodes leading to a catastrophic release of radioactivity
1986	The world's largest 14,000 MW Itaipu Dam, along the border of Brazil and Paraguay, is opened
1986	NOVA, an experimental laser fusion device at the Lawrence Livermore National Laboratory create the first laser fusion reaction. Fusion, however, remains elusive through the end of the century
1989	Exxon Valdez, a 1,260,000 barrels-of-oil tanker strikes a reef in Alaska's Prince William Sound, spilling an estimated 240,000 barrels of crude oil
1992	The biggest array of 9600 thin film photovoltaic modules ever assembled starts operation in Davis, California, delivering up to 479 kW, enough for over 100 homes
2000s	Computerization with Internet information exchange, sophistication and globalization (see Energy Future Outlook section above)

Note: All early dates are approximate and other dates may differ due to reference to either conception, patenting, publication, or first application, in different sources[8–10] or elsewhere.

Glossary

Energy: It is a fundamental property of a physical system and refers to its potential to maintain a system identity or structure and to influence changes with other systems (via forced–displacement interactions) by imparting work (forced directional displacement) or heat (forced chaotic displacement/motion of a system's molecular or related structures). Energy exists in many forms: electromagnetic (including light), electrical, magnetic, nuclear, chemical, thermal, and mechanical (including kinetic, elastic, gravitational, and sound).

Energy Conservation: It may refer to the fundamental law of nature that energy and mass are conserved—i.e., cannot be created or destroyed, but only transferred from one form or one system to another. Another meaning of energy conservation is improvement of efficiency of energy processes so that they could be accomplished with minimal use of energy sources and minimal impact on the environment.

Energy Conversion: A process of transformation of one form of energy to another, such as conversion of chemical to thermal energy during combustion of fuels, or thermal to mechanical energy using heat engines, etc.

Energy Efficiency: Ratio between useful (or minimally necessary) energy to complete a process and the actual energy used to accomplish that process. Efficiency may also be defined as the ratio between energy used in an ideal energy-consuming process vs energy used in the corresponding real process, or vice versa for an energy-producing process. Energy, as per the conservation law, cannot be lost (destroyed), but the part of energy input that is not converted into useful energy is customarily referred to as energy loss.

Industrial Revolution: Controversial term referring to the development of heat engines and use of fossil fuels, first in Britain and later in other countries, from 1760 to 1850.

Nonrenewable Energy Sources: The energy sources (such as fossil and nuclear fuels) created and accumulated over a very long period in the past, for which the creation rate is many orders of magnitude smaller than the consumption rate, so that they will be depleted in a finite time period at the current rate of consumption.

Renewable Energy Sources: The continuously or frequently available (renewed daily or at least annually) energy sources—solar energy, wind, water flows, ocean and tidal waves, biomass, and so on—that, for all practical purposes, are expected to be available forever.

REFERENCES

1. Kostic, M. Work, power, and energy. In *Encyclopedia of Energy*, Cleveland, C.J., Ed.; Elsevier Inc.: Oxford, U.K., 2004; Vol. 6, 527–538.
2. Kostic M. Irreversibility and reversible heat transfer: The quest and nature of energy and entropy. In *IMECE2004, ASME Proceedings*. ASME : New York, 2004.
3. Energy Information Administration. *Annual Energy Review 2004*. U.S. Department of Energy, Available at: www.eia.doe.gov/aer (accessed on June 2005).
4. Energy Information Administration. *International Energy Annual 2003*. U.S. Department of Energy, Available at: www.eia.doe.gov/emeu/iea/wec.html (accessed on June 2005).
5. Energy Information Administration. *International Energy Outlook 2005*. U.S. Department of Energy. Available at: www.eia.doe.gov/oiaf/ieo (accessed on June 2005).
6. Energy Information Administration. *Forecasting and Analysis of Energy*. U.S. Department of Energy. Available at: www.eia.doe.gov/oiaf/forecasting.html (accessed on June 2005).
7. Lawrence Livermore National Laboratory. *The Energy and Environment Directorate*. Available at: http://eed.llnl.gov (accessed on June 2005).
8. Smil, V. *Energy in World History*; Westview Press, Inc.: Boulder, CO, 1994.
9. California Energy Commission. *Energy Time Machine*, Available at: www.energyquest.ca.gov/time_machine/ (accessed on June 2005).
10. Bejan, A. *Advanced Engineering Thermodynamics*; Wiley: New York, 1988.

Enterprise Energy Management Systems

Gregory Cmar
Bill Gnerre
Kevin Fuller
Interval Data Systems, Inc., Watertown, Massachusetts, U.S.A.

Abstract
Building automation systems and metering systems are an enormous source of valuable operational data. However, it is rarely put to use, as facilities professionals do not have adequate means to collect and manage the data. The information is lost and decisions are made without solid operational facts, resulting in underperforming buildings and physical plants. An Enterprise Energy Management System (EEMS) addresses this by collecting all of the operational data from building systems, organizing it, and providing users with highly interactive access to the infomation contained within the raw data.

This article highlights the key functional requirements of an EEMS, and focuses on the complexities of collecting data from the many disparate sources and making it a highly valuable information source for operators, engineers, technicians, and management.

INTRODUCTION

It is all about the data. Current building automation systems (BAS) and metering systems produce huge amounts of valuable data—most of which is thrown away. Facilities staff and engineers typically use as little as 1% of the available data for making decisions about operations, building and plant optimization, commissioning, design engineering, performance contracting, automated diagnostics, and other related tasks. This is the fundamental role of an enterprise energy management system (EEMS)—to collect all of the available data related to energy consumption and costs, organize it, and provide users with access to the engineering facts that live within the data.

Unlike control systems and diagnostic tools, an EEMS provides a complete view of an entire facility's operations, not just component parts. This level of insight is the only way to achieve all of the business benefits possible, including lower energy consumption, improved space comfort, extended equipment life, better use of staff, energy purchase savings, demand reduction, better engineering designs, lower construction costs, and much more.

This entry highlights the key functional requirements of an EEMS and focuses on the complexities of collecting data from the many disparate sources, making it a highly valuable information source for operators, engineers, technicians, and management.

Keywords: Energy management; Facilities operations; Chiller plant; Boiler plant; Optimization; HVAC; Interval data; Utility billing.

EEMS Defined

An EEMS consolidates energy-related data (sources, costs, and control and monitoring points) into a data warehouse and provides a platform for tools to access the data easily and obtain actionable information.

The EEMS makes data-driven information available so that the end user is able to perform in-depth diagnostics, engineering analysis, and monitoring in a small fraction of the time it took with earlier methods. By publishing information for consumption outside the traditional facilities management area, an EEMS supplies critical information to department heads, financial staff, and executive management as well.

Five simple but crucial principles form the basis of an ideal EEMS:

1. All energy-related data is consolidated into a centralized data warehouse.
2. The collected data is "normalized" and structured.
3. Access to data is interactive to facilitate distillation of actionable information.
4. The system makes it easy to measure and verify results.
5. The system provides a platform that embraces industry standards for data collection, management, analysis, and publication.[1]

It is worth noting what an EEMS is not: it is not a control system, and it should not be confused with a BAS. An EEMS is much broader in scope. It provides a platform for data collection, data access, diagnostic and monitoring capabilities, a historical data warehouse, and a lot more, as detailed throughout this entry. Similarly, an EEMS is not a utility billing system. It encompasses billing and meter data but extends beyond by connecting billing information directly to the related operational data.

EEMS PRINCIPLE 1: CONSOLIDATE ALL ENERGY-RELATED DATA INTO A DATA WAREHOUSE

Energy-related data comes from purchased utilities, generated utilities, BAS, metering systems (both advanced and manually read), weather, computerized maintenance management systems (CMMS), and space planning systems. Additionally, an EEMS manages rate and billing data, users, and organizational information.

To be able to utilize energy data, the first step is to identify and collect the right data into a data warehouse so that accurate and actionable information can be available. The EEMS needs to collect all data, as one cannot optimize the whole system by optimizing each component.

This section identifies the different data sources and attributes that define an EEMS and populate its data warehouse.

Purchased Utilities

Most facilities purchase electricity, fuel (usually natural gas or oil), and water. Some also buy chilled water and steam. An EEMS manages both consumption and billing information, and presents this data in an intelligible, clear, and actionable format. Fig. 1 is a simplified view of the issues related to collecting utility data.

Consumption information is typically time based, regardless of the type of utility. For example, electric utilities commonly use 15-min intervals; natural-gas utilities use daily intervals; and water uses monthly or quarterly intervals. Eventually, these different time series need to be "normalized" so that information can be presented in consistent intervals.

Billing information is equally complex. Large facilities often have multiple vendors for each utility type, each with differing rates, billing cycles, and pricing structures, resulting in a significant amount of billing data to collect and manage. For example, a university may have 1000–10,000 utility bills per year.

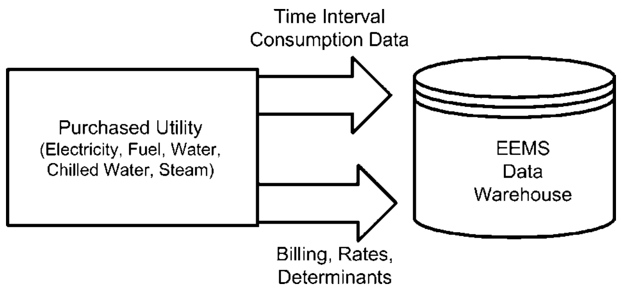


Fig. 1 Billing and consumption data should be collected in tandem.

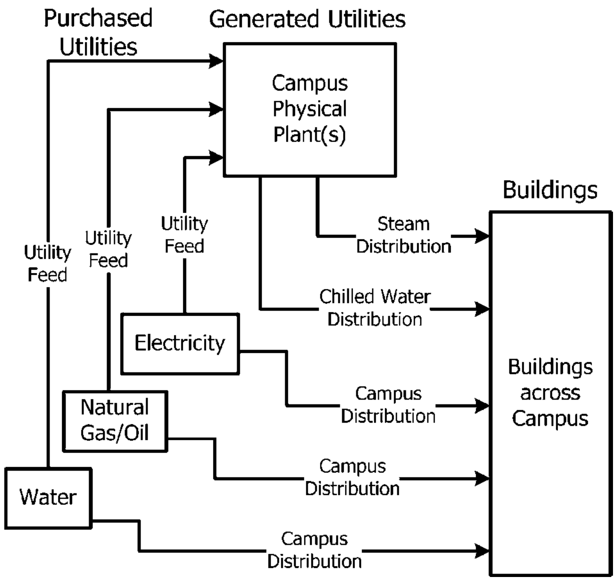


Fig. 2 Institutions typically deliver purchased and generated utilities to the facility.

Generated Utilities

Most large facilities have their own physical plants that generate and distribute chilled water and steam. An EEMS collects data from the control system(s) of the plant, as well as from the distribution network and buildings (Fig. 2).

The EEMS also collects billing information, just as it does for purchased utilities. Facilities that generate their own chilled water, steam, or even electricity will have their own rate structures and determinants, and they will bill internally based on consumption.

Building Automation Systems

Many institutions operate more than one BAS, so an EEMS provides a complete, holistic view by collecting data from all systems and overcoming the limitations of relying on the BAS for data. It is this systemwide view that enables organizations to diagnose operations and energy usage more quickly and effectively (Fig. 3).

Extracting Data from the BAS

To provide a comprehensive picture, an EEMS requires data from all control and monitoring points. Gathering data only from monitoring points (without control data) means that while you may identify something to improve,

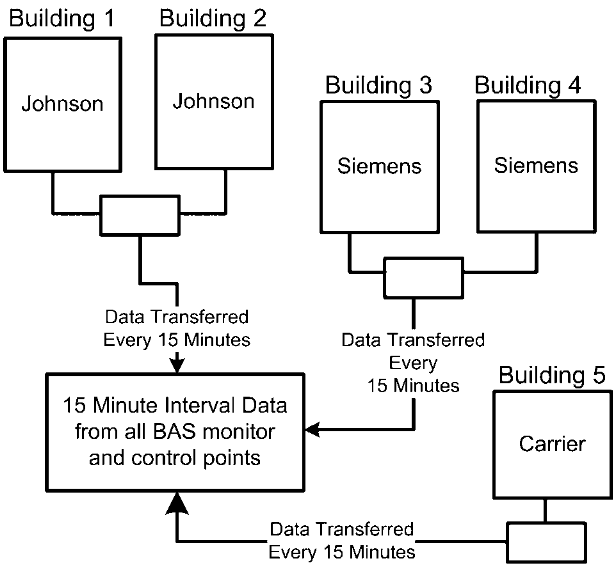


Fig. 3 Energy management system (EEMS) collecting data from multiple building automation systems (BAS).

it is unlikely that you'll gather sufficient information to know with certainty how to improve it (Fig. 4). (This is what happens when one installs advanced metering systems instead of an EEMS.)

At large facilities, it is not unusual to have 30,000–130,000 points (or more), each generating one value every 15 min. For every 30,000 points, there are over a billion data intervals per year. Without an EEMS, much, if not all, of this information is thrown away (Fig. 5).

Meters and Metering Systems

Meters and metering systems are typically located at buildings and throughout the distribution system to measure the use of electricity, steam, chilled water, fuel, etc. (Fig. 6).

For electricity, access to metering data via the metering database is a relatively straightforward process, but for those utilities that do not leverage advanced metering systems, data can be obtained through the BAS (Fig. 7).

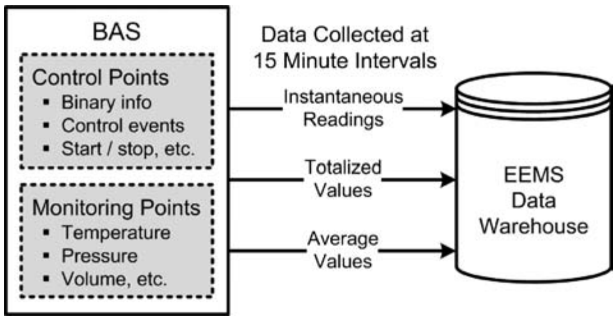


Fig. 4 The energy management system (EEMS) collects various kinds of data from control and monitoring points every 15 min.

In addition to the automated data collection described above, an EEMS will incorporate manually collected data.

Other Data Sources

Weather has a large impact on energy consumption and forecasting. For the most accurate results, an EEMS includes official weather data from a local airport, in addition to local sensors that may be inaccurate for various reasons, such as being located too close to a building, in the sun, broken, etc. (Fig. 8).

Space planning data is used by an EEMS to identify energy costs at the space level. Space planning systems contain information concerning the use and allocation of all areas within a facility. They map the hierarchy of the facility and understand the relationships between space and cost centers.

Market and pricing data affects the ability to purchase energy at favorable rates. Prices, pricing structures, and regulations constantly change. The role of the EEMS in purchasing energy is threefold:

- To display past and future energy usage patterns
- To convert and present utility billing and consumption usage into an equivalent real-time price
- To present real-time pricing information.

Easy access to actual and predicted usage patterns enables the organization to make more-informed utility purchasing decisions than ever before.

Do Not Throw Data Away

Large organizations spend millions of dollars annually on energy, millions on systems like BAS, and tens of millions for heating, ventilation, and air conditioning equipment, yet they throw away most of the data these systems generate, largely because there is so much data, and they are difficult to access, manipulate, and interpret.

An EEMS captures these data and leverages them to provide a complete picture of utility consumption and an organization's energy infrastructure over time so that the most expeditious analysis can take place to reduce the total cost of ownership and operation.

EEMS PRINCIPLE 2: NORMALIZE AND STRUCTURE DATA

Energy management system (EEMS) Principle 1 focuses on the importance and requirements of getting data into the system. Principle 2 addresses constructing a data warehouse based upon the data least common denominator: a standard time interval.

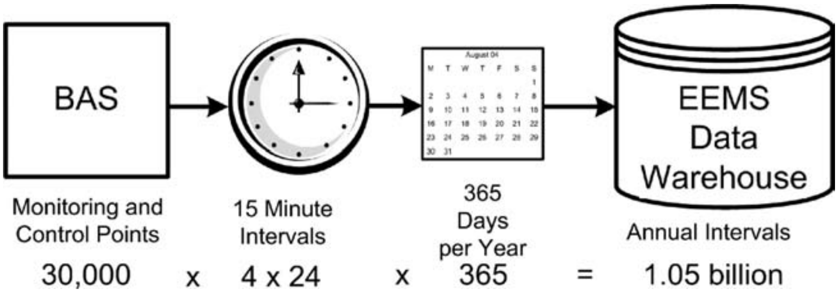


Fig. 5 The ability of an energy management system (EEMS) to store, manipulate, and display very large volumes of data in an efficient manner is mission critical.

Optimal Time Interval

An EEMS uses a standard time series interval, selected to ensure sufficient data to identify transitions. This means that there should be enough data points gathered to discern performance fluctuations across transition time periods (between day and night, office hours and nonworking hours, etc.) so that behavior patterns and problems become apparent quickly.

For this reason, and because electricity is frequently metered within the same time interval, 15 minutes is an appropriate time for EEMS normalization. Longer intervals do not provide sufficient data granularity to show all behavioral changes. Shorter intervals can be used, but one has to manage the trade-off between more data and more information. You also need to consider how long data will be kept—ideally, for the lifetime of a piece of equipment (20 years or more).

Normalized Interval Data

Normalizing means that data from all sources is stored in the warehouse in the same time interval. Not all data sources provide data in the same interval, so the EEMS reconciles these differences.

Interval data inconsistencies are a fact that must be acknowledged and addressed. A well-implemented EEMS accepts these inconsistencies and fills the gaps with estimates that, when totaled, account for the total energy consumed and represent the pattern of that energy usage in a precise and accurate manner. For meter data, the relevant utilities' processes for validating, editing, and estimating (VEE) should be followed.

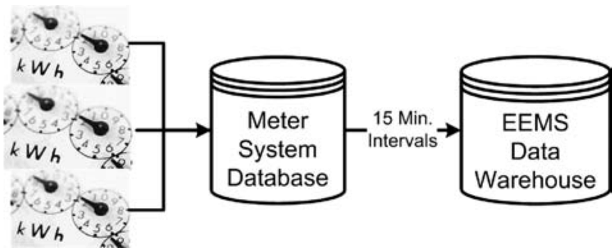


Fig. 6 A connection between the metering databases and energy management system (EEMS) must exist to synchronize data views of meter and building automation systems (BAS) data.

Data Warehouse Structure and Hierarchy

To derive value from such a large amount of data, a defined structure and hierarchy must exist to make the data readily consumable. The EEMS data structure requires flexibility too, because physical configurations are constantly changing—buildings are added, equipment fails unexpectedly, organizational changes lead to space modification, etc.

Warehouse Objects

The data structures in an EEMS parallel the physical facility, allowing it to be easily used to focus on an individual building or piece of equipment.

An EEMS warehouse supports logical objects—sets of information and relationship hierarchies—that allow for this structure. Table 1 shows six such warehouse objects and the hierarchy that exists within them. The warehouse structure also supports the way different elements within the object inter-relate—for example, the way meter interval data connect to billing rate data and physical building data.

EEMS PRINCIPLE 3: PROVIDE INTERACTIVE ACCESS TO ACTIONABLE INFORMATION

Defining Interactive and Actionable

Interactive access allows the users to work with the data in a dynamic fashion, moving seamlessly through the data with tools that provide near-instantaneous response. This allows users to "work the way they think" rather than being limited to a series of static queries and reports.

Actionable information is usable as the basis and rationale for effective decision-making. To create

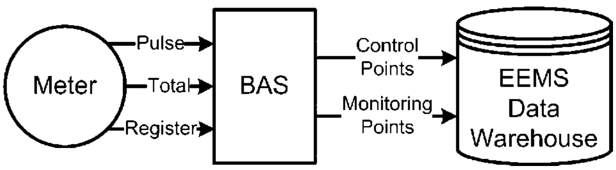


Fig. 7 Data can be obtained and transferred through a building automation systems (BAS).

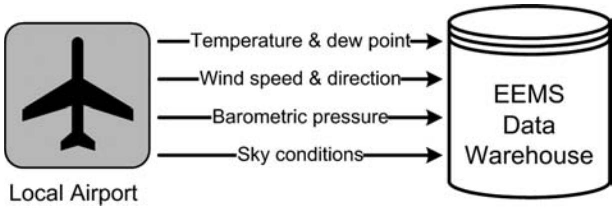


Fig. 8 METAR weather data should be used within an energy management system (EEMS).

actionable data, the EEMS should collect both monitoring and control points. With monitoring data alone (without control information), it is difficult to verify or quantify the savings opportunity.

Table 1 Object hierarchies maintained by an energy management system (EEMS) data warehouse

Facility objects
Site
Zones
Buildings
Floors
Rooms
Higher education general information survey (HEGIS) group
HEGIS classification
Issue objects
Issues
Organizational objects
Departments
User objects
User privileges
Users
Interval data collection objects
Building automation systems (BAS) and OLE for process control (OPC) point data
Calculations
Balance efficiency calculation
Cooling efficiency
Cost calculations
Delta T
Theoretical water loss
Etc.
Weather data
Meter objects
Utility type
Utility information
Account information
Meter information
Rate information
Billing information

Time and Usability Matters

With gigabytes or even terabytes of data to manage, the use and operation of an EEMS cannot be an arduous and time-consuming task. An EEMS accesses and presents data to users at the "speed of thought," enabling users to view information created from hundreds of thousands of data intervals in a few minutes. If users have to wait for this information, their thoughts will wander. If they continually have to wait, and if they are not constantly engaged, their reaction will be that the EEMS is wasting their time.

Currently, most facility personnel waste an enormous amount of time collecting and distributing data. This wasted time is culturally accepted as part of the job, when in fact, instead of spending days manually gathering and piecing together data, a well-implemented EEMS can deliver it in seconds. With immediate access to the data and information, staff frees up time to devote to the real engineering work of diagnostics, analysis, and planning.

The time required to gain insight is directly related to the usability of the system. It is essential that data can be assembled dynamically by the user via a simple, intuitive interface. The elements of the interface include the data organization and visualization, and the program interface itself. A well-integrated EEMS supports other aspects of the facility staff's workflow, such as tracking identified issues or interacting with external analysis tools.

Data Organization and Presentation

An EEMS categorizes data in meaningful ways, such as organizing within the facility hierarchy (zones, buildings, floors, rooms), organization (departments, rooms), or systems (chiller system, air handling system). In most cases, data access occurs through multiple views so that, for example, a building manager can examine the building information while a plant engineer can look at chiller operations facility-wide. Users can "drive through the facility" from their desktops by viewing the information. An EEMS also provides a method for users to define their own organized views.

The Importance of Trend Data

Trend data is the foundation of diagnostics, monitoring, measurement, and verification, and building a historical record of facility operations. While BAS typically contain a function called trend logs, they are a poor mechanism to collect trend data for the EEMS due to a variety of limitations. Trend lines from the EEMS interval data provide a data visualization and interaction paradigm that does not suffer these same limitations (see Table 2). The differences elevate, by orders of magnitude, the

Table 2 Some of the uses, users, and differences between building automation systems (BAS) trend logs and energy management system (EEMS) trend lines

Application of BAS trend logs and EEMS trend lines

Criteria	**BAS trend logs**	**EEMS trend lines**
When to use		
Diagnosing operations	After problem has been identified as under the control of the BAS and further data is needed for final diagnosis	Always—superior tool for nearly all diagnostics and all cases where historical data or data outside the BAS must be considered
Monitoring operations	When real-time data for a small number of BAS points needs to be watched	Always—provides the ability to monitor hundreds of trends in minutes, combining data from any and all sources
Typically used by	BAS control engineers and technicians	BAS control engineers and technicians, energy engineers, area mechanics, facility managers, performance monitoring contractors, commissioning agents, HVAC design engineers
Technology perspective		
Data storage	Stores point data for trends defined	Stores data for all points from all systems
Data availability	Only those points explicitly trended, with the total number of trends extremely limited (a few points)	Historical trend data always available for every point in the system (many thousands) without affecting control system performance
Time interval of data	Captures data in increments from milliseconds to minutes	Captures data from all systems at regular intervals (e.g., 15 min)
Displays data from	Native BAS	Multiple BASs, metering systems, utilities, weather, billing
Display time period	Typically a few days or weeks	Between a day and a year, with historical data going back years
Data storage	Up to a few months and data is discarded	Up to 20 years

effectiveness of an EEMS for diagnostics, monitoring, and other applications, which results in actionable information.

Building automation systems trend logs do have their place in providing needed information—where their ability to collect real-time data is useful—as shown in Table 2.

Calculated Data: Providing Actionable Information

The efficiency gains in having commonly desired calculations available for monitoring and diagnostics are tremendous, and they have a dramatic impact on usability. A visual display of an ongoing trend built on a calculation can provide insight instantly that would otherwise take hours of analysis in Microsoft Excel.

More complicated calculations can provide users an overall operational efficiency rating, the power consumptions, and the energy cost for each 15-min period. Calculations of this complexity rely on the EEMS's ability to integrate fully data from all sources and present them in a normalized fashion as actionable information.

A sample of calculations includes

- Balance equations
- Chiller efficiency
- Chiller total cost
- Chiller plant total hourly cost of operations
- Cooling tower cost and efficiency
- Cooling tower makeup water cost and efficiency
- Delta T
- Pump brake horsepower
- Pump efficiency
- Pump kW
- Theoretical water loss
- Tons output.

Application Interface

All the usability factors mentioned must come together in the software user interface. It is how end users interact with the data—through mouse clicks, menu selections, expandable data trees, contextual menus, dragging and dropping, etc.

The EEMS allows users to take an iterative approach with each action, building on the last for diagnostic purposes. Monitoring is fast and efficient, allowing users to cycle quickly through hundreds of trend lines in minutes. Ultimately, it is the combination of information display (trend lines, calculations, etc.), user interface, and system performance of the EEMS that allows the user to work at the speed of thought and to improve the quantity and quality of decisions made.

Present User-Specific Information

An EEMS supports a variety of energy-related applications and users, including energy engineers, HVAC design engineers, technicians, area/building mechanics, facilities engineers and managers, commissioning agents, energy purchasers, and performance contractors.

An EEMS provides a variety of output options that enable everything from detailed reports with data tables and charts of multiple trend lines to summary reports that roll up information into cost breakdowns and overall operating-efficiency ratings.

Users of information from an EEMS extend well beyond facilities and maintenance staff to building managers, department heads, finance personnel, executives, and managers—anyone who has occasional needs for some facilities information. These "tier two" users typically do not need interactive access to the data at the individual point level and are well served by predefined reports focused on business issues, not engineering details. Enterprise-level energy and operational data published to meet their specific needs and interests will help realize the many business benefits listed later in this entry.

EEMS PRINCIPLE 4: MEASURE AND VERIFY RESULTS

All too often, performance measurement and verification (M&V) ends up neither measuring nor verifying performance. It is a simple case of not having access to the data to do M&V properly—a problem an EEMS solves. Without an EEMS, verifying and measuring results of optimization efforts are largely left to engineering estimates and educated guesswork. At the very least, it is an extremely time-consuming process to gather thorough operational data. The ability to quantify improvements accurately and monitor their persistence gives facility directors the tools to hold contractors and their own staff accountable, and these tools can help directors increase their credibility with upper management.

In measuring and verifying results, it is important to define terms often used to justify and quantify the impact of investments in utility operations.

Energy Savings vs Dollar Savings

An EEMS accounts for both energy savings and cost savings (actual dollars). Just because the utility bill dropped does not necessarily indicate that money/energy has been saved; rather, it confirms only that less money was spent. It verifies nothing about reduced energy consumption.

Factors like price, weather, new construction coming online, and consumption rates are required to understand whether actual energy savings have occurred. To realize true energy savings, which in turn lead to dollar savings, consumption must be reduced independent of these factors, and this can be achieved only when energy usage is being controlled more efficiently.

Stipulated Savings vs Real Savings

Stipulated savings typically are savings amounts specified and agreed to up front with an entity assisting with funding energy efficiency improvements—for example, a performance contract for retrofitting lighting that uses 25% fewer watts or replacing an older chiller with a newer, more efficient model. Stipulated savings often do not translate into real savings because more variables than just the efficiency of the equipment come into play.

The stipulated amount will not take into account issues such as the percentage of lighting fixtures that aren't working prior to the retrofit, or behavioral changes in hours the new lighting is in use or the percentage of lights turned on, and often, this amount will not consider the additional heating energy required during winter months because of the lower wattages. This isn't to say that the efficiency gains promised don't exist, but that one may not see the expected consumption reductions in utility bills. An EEMS allows one to view both the data for the electricity consumed by equipment type and for the utility consumed to heat (or cool) the area, all within the context of a specific space, allowing a real savings assessment to be made.

Use Life-Cycle Costing

An important concept to adopt is life-cycle costing. It is the most appropriate way to assess equipment and building costs. The initial purchase of HVAC equipment is a significant capital investment, but its true costs lie in this number plus the cost of its operation, service, maintenance, and total life span.

EEMS PRINCIPLE 5: A PLATFORM THAT EMBRACES INDUSTRY STANDARDS

Using industry standards and an open architecture is the right way to build any enterprise-class application. This has been proved repeatedly at all levels of technology and business, where broad support and interoperability are significant benefits. A platform architecture and the use of standards protect the organization by minimizing dependency on any single vendor, even allowing functionality to be added outside the vendor's development cycle.

There are standards in several areas that an EEMS should adhere to.

Operating system. There are three platforms, sufficiently open and standardized, that an EEMS could run on. The first, and by far the most popular, is Microsoft Windows. It offers the greatest availability of tools and options, and is already installed and supported nearly everywhere. The Microsoft.NET platform is excellent for developing and integrating application components. Other options include Linux, which has strong server support and tools such as J2EE but is a limited end-user platform, and a Web-based solution, which involves Windows or Linux servers using Web browsers for the application front end (which instills limitations on the user interface).

Database management. Data storage for an EEMS should be open, accessible, and interoperable with other systems using a standard relational database (e.g., SQL Server, Oracle, etc.). Proprietary data managers will handcuff users to rely on the vendor for everything.

Data collection. As discussed earlier, data are the lifeblood of an EEMS, and access to these data is a complex and arduous task. Standards such as OLE for process control (OPC), BACnet, and LonWorks enable cost-efficient solutions for data collection.

Analysis. Data analysis tools range from the most general and broadly available (Excel) to highly specialized analytics. Ideally, the EEMS will provide direct support for Excel, allowing users to take advantage of analysis routines already developed. Minimally, the EEMS will provide data export into any analysis program.

Space planning and classification. For colleges and universities, the EEMS should support higher education general information survey (HEGIS) groups and classifications for space planning. It should also interface with market-leading space planning systems.

Publishing. Making information available throughout the organization requires a standard format, such as HTML. For long or data-intensive publishing, other formats may be appropriate, such as PDF, XML, or Microsoft Office.

Geographically Dispersed Enterprises

Not all enterprises are located in a single building or campus. When implementing an EEMS across a series of locations, it is preferable to collect data locally at each facility, primarily to minimize the number of possible points of failure that could interrupt data collection. A well-architected EEMS will allow centralized users to view and combine data from the individual locations, thereby still realizing the full benefits of an enterprisewide understanding of operations.

The size of the individual facilities may influence the local/central decision. For example, a school district may choose to deploy an EEMS centrally because the engineering and operations staff may be a single, district-wide group. However, local technicians and mechanical staff should have access to the data pertaining to their school.

BUSINESS APPLICATIONS OF AN EEMS

The EEMS has many business applications, each with different benefits to the facility. Some of these applications deliver benefits that are operational savings (reduced energy consumption); some enhance the infrastructure; and others address the business of facilities operations. The most common applications of an EEMS are:

- Operational diagnostics and monitoring
- Efficient building control strategies
- Building and physical plant commissioning (existing buildings and new construction)
- Validating engineering design/redesign and construction
- Chiller plant efficiency calculations
- Controlling building comfort
- Enhanced customer service
- Accurate energy cost allocation
- Capital-request justification
- Information publishing
- More accurate budgeting and forecasting
- Increased accountability and credibility
- Purchased utility accounting
- Vastly improved performance M&V.[2]

Financial Benefits of EEMS

It is clear that an EEMS is able to provide the infrastructure to support many different business purposes. It is important that these applications of EEMS technology are also able to deliver a rapid return on investment in a number of areas, including:

- Lower operational costs
- Positive cash flow
- More effective staff deployment
- Greater indirect cost recovery
- Reduction in equipment maintenance costs and increased equipment life
- Improved efficiency of energy purchasing.

CONCLUSIONS

This entry provides the reader a definition, a functional description, and insight into the complexities of an EEMS. For readers considering an EEMS purchase, this entry does not address in detail issues such as the functionality/value/cost relationships. Larger facilities—those with a million square feet of space or more—should look for an EEMS that is complete and meets the ideal criteria. At this writing (2006), these systems have not scaled down in size sufficiently to necessarily be cost effective for smaller facilities. This will happen over time, especially as BAS vendors make data more easily available.

An ideal EEMS will adhere to the following criteria:

- It collects data from all monitoring and control points from all sources, gathering consumption data, control settings, and billing and rate information to provide a holistic view.
- It normalizes and structures the data within a data warehouse, providing flexibility to handle complex relationships, hierarchies, calculations, and adjustments to meet evolving requirements.
- It presents actionable information through trend lines and other reporting/graphical formats. Performance and user interface must combine to provide interactive access to the data.
- It delivers value and savings that can be measured and verified.
- It is more than an application—it is a platform for energy management and other facilities operations that is expandable and conforms to open standards.

An effectively implemented EEMS provides unparalleled insight into facility operations. It nearly eliminates the time wasted by staff gathering needed data. It provides individuals the ability to find and address inefficiencies rapidly (fixing the root cause, not just treating symptoms) that can result in immediate cost savings and an ongoing financial return—often when these problems have gone undetected for many months or even years.

An EEMS presents the opportunity to span all existing BAS and energy-related data so that assessments can be made in the context of the whole facility, environment, and billing climate. Complete information leads to better decisions—decisions that address building comfort, energy consumption, operational costs, capital investments, and stewardship of the assets.

REFERENCES

1. Cmar, G.; Gnerre, W. Defining the next generation enterprise energy management system. In *Web-Based Energy Information and Control Systems: Case Studies and Applications*; Capehart, B., Capehart, L., Eds.; Fairmont Press: Atlanta, GA, 2005; 403–434 (Chapter 32).
2. Interval Data Systems, Inc. *Innovating the Business of Facilities Operations: A Framework for the Next Major Advancement in Facilities Operations*; Interval Data Systems, Inc.: Watertown, MA, 2005.

Environmental Policy

Sanford V. Berg
Director of Water Studies, Public Utility Research Center, University of Florida, Gainesville, Florida, U.S.A.

Abstract
Environmental policy is designed to reduce the negative impacts of economic activity on water, air, and other natural resources. It is difficult to identify and quantify the damage caused by pollution on human health or inflicted upon sensitive ecosystems, so determining the benefits and costs of remediation policies is often extremely difficult. Policy-makers respond to domestic political pressures by devising institutions and instruments to address pollution and environmental sustainability. Cost-benefit analysis (CBA) and cost-effective analysis provide frameworks for systematically identifying and evaluating abatement strategies that avoid potential inefficiencies or inequities.

INTRODUCTION

When economic activity leads to pollution and over-use of common property resources, government intervention can improve social welfare. Pollution involves a market failure in which damages caused by a producer or consumer are imposed on third parties. These damages can involve personal health, the physical deterioration of buildings, and foregone options for the future. Of course, if transaction costs are low, those that are causing pollution damage can be taken to court if the liability rules are clear. Destruction of common property resources such as losing unique ecological habitats, endangering particular species, or destroying valued scenic vistas is another form of market failure affecting the environment. Because there may not be clear property rights to such elements of the environment, these common property resources can be overutilized. Given the lack of well-defined property rights, government enacts environmental laws to address these market failures. However, identifying and quantifying the damage caused by pollution sources or inflicted upon sensitive ecosystems can be difficult, making any determination of the benefits and costs a contentious exercise. Consequently, choosing policies that define the extent of the environmental protection can be both contentious and problematic.

The next section describes the multidisciplinary inputs that are incorporated into environmental policy analysis, selection, and implementation. Other topics addressed here include policy impacts, the burden of proof, economic evaluation, and the strengths and limitations of policy options.

Keywords: Environmental impacts; Regulation; Cost-benefit analysis; Cost-effective analysis.

MULTIDISCIPLINARY APPROACH TO DEVELOPING ENVIRONMENTAL POLICY

Environmental economics is the study of how economic and environmental issues interact. Issues addressed by environmental economists include but are not limited to evaluating ways to reduce pollution, analyzing the trade-offs between using renewable and nonrenewable resources, or estimating monetary values for ecosystems or habitat. While no single field of study contains all the insights needed to develop and implement sound environmental policies, the focus here will be on economics because it provides a system for incorporating many perspectives and it is the framework by which environmental policy is designed and evaluated. Depending on the burden of proof, the resulting policies might be excessively stringent (costly relative to their benefits) or inadequate for the protection and preservation of environmental features that affect human health and welfare and have intrinsic value.

We know from materials balance that human activity does not create matter but only changes its form, concentration, and location, thus there is a need for physical sciences such as chemistry, physics, and biology to help inform environmental policy. While all societies affect natural systems, the scale of potential impacts has grown with economic development. There is evidence that as incomes rise, citizens are willing to devote relatively more resources to controlling environmental impacts. Moreover, many citizens would like to see much more attention given to reducing current damages and limiting the risks for future harm, hence an understanding of societal and political dynamics is also important for informing environmental policy.

The development and implementation of sound environmental policy draws upon information and procedures from many fields of study. Here, economics is utilized as the framework for integrating the concepts, measurements, and values required for the steps:

1. Determine appropriate regulatory objectives (through citizen participation in political processes and community consensus-building).
2. Balance those objectives to determine regulatory priorities.
3. Identify and legislate oversight responsibilities for environmental agencies.
4. Develop (a) mechanisms for monitoring environmental impacts (such as ambient air and water quality) and (b) methodologies for integrating new scientific understandings of environmental impacts into the policy prioritization process.
5. Define the appropriate targets for different types of pollutants and the protection of biodiversity.
6. Determine (and then apply) the appropriate policies for meeting objectives.
7. Analyze environmental indictors on a regular basis, checking for noncompliance.
8. Evaluate the impacts, recognizing potential biases in the measures and the ways impacts are valued.
9. Establish an effective process for monitoring and reviewing the framework, including the penalties and sanctions applied when there is noncompliance.

These steps require input from a number of disciplines that shape the way we see things. Although technical training allows analysts to delve deeply into subjects in a consistent manner, awareness of other disciplines' perspectives can be important for constructive environmental policy-making, including:

- Engineers look to technology for solutions to environmental problems. They are able to incorporate new (often expensive) control technologies into energy extraction, production, consumption (energy efficiency), and pollutant disposal and storage (as with nuclear waste).
- Meteorologists and hydrologists analyze pollution transport in air and water systems. They have a deep understanding of the impacts of discharges under different conditions. In conjunction with demographers and epidemiologists, they can estimate the doses received by different population groups.
- Medical scientists and toxicologists analyze the dose-response relationships for citizen health, conducting exposure and risk assessments.
- Ecologists study the impacts of pollutants on the local and global environment, assess the value of ecosystem services, and track invasive species and biodiversity. Climate scientists help assess the causes and consequences of changes in local and global temperatures and other weather patterns.
- Materials scientists look at damages caused by air and water pollution. The associated impacts include cleaning and painting buildings, treatment costs, and shorter life spans for affected equipment.
- Political scientists focus on issues of power, legitimacy, social cohesion, and the roles of different stakeholder groups in influencing environmental policies. Consensus is critical because ultimately, in a democratic system, there needs to be widespread agreement on the desired outcomes if the system is to avoid instability.
- Economists emphasize the importance of efficiency in resource allocation. They apply benefit-cost analysis and tend to depend on price signals to provide incentives for the adoption of appropriate control technologies and conservation measures.
- Planners deal with land-use and zoning issues, given population growth projections. Planners integrate legal constraints with historical experience, bringing topological, aesthetic, and geographical elements to the analysis.
- Archeologists and anthropologists provide insights on the impacts of dams, mines, and their related economic activities on unique historical sites, local populations, and indigenous groups. Such impacts create difficult valuation issues.[1,2]
- Lawyers spotlight the institutions of policy implementation. For example, rules and regulations attempt to pay significant attention to procedural fairness. Due process contributes to the legitimacy of outcomes. If different parties perceive that there is no transparency and no opportunity for participation, environmental policy will be perceived as unreasonable and the laws will either be changed or they will be disobeyed in a variety of ways.
- Environmentalists advocate sustainability and environmental equity. The by-products of energy production affect public health and have environmental outcomes. Those impacts have economic value, but often that value is nonmonetary or difficult to quantify. For example, generation and transmission siting decisions incorporate impacts on biodiversity and sustainability.
- Ethicists help society understand personal values and notions of stewardship. Humans have a clear responsibility to leave future generations with a legacy of sound institutions and a clean environment, though the best means to this end are often not obvious.

Thus, physical, biological, and social scientists attempt to uncover patterns and identify lessons to help us improve policy. Given the complexity of environmental issues, most environmental problems are managed, not solved.

IMPACTS

Energy production and consumption impact people and the environment in a number of ways. For example, activities can damage ecosystems in the extraction phase (oil drilling or coal mining) or involve cross-media emissions in the consumption phase that can lead to further ecosystem damage. Emissions can be from a single point or a mobile source. In addition, they can be continuous or intermittent (with exposure and impacts depending on wind and other weather conditions or the presence of other chemicals). The transport mechanism can be complicated and involve multiple jurisdictions (as with SO_2 and NO_x—emissions lead to “acid rain” or ozone problems in downwind areas).

Air

Issues range from local concentrations of particulate matter in the atmosphere to concerns over anthropogenic climate change. Consequences for health, ecosystems, agriculture, coastal settlements, species survival, and other impacts make atmospheric change a serious policy issue. For example, long-range transport means pollutants cross national boundaries and require coordination. Other pollutants—such as greenhouse gas emissions of CO_2—require coordination not due to transport, but because the effects are global in nature regardless of where emissions occur.

Water

Effects of contaminants vary in surface waters and groundwater. The United States has primary standards to protect public health (with maximum contamination levels [MCLs] for toxic wastes). Secondary standards and associated MCLs are meant to protect public welfare (for example, ensuring that the taste, odor, and appearance of groundwater do not result in persons discontinuing water use). Other environmental issues include species loss and dealing with nonindigenous, invasive species.

Land Use

Siting is an issue for electricity generators, transmission lines, and distribution systems (other aspects of land use include urban sprawl and availability of land for agriculture. The focus here is on the environmental impact of energy systems. For example, social investments in mass transit affect emissions from mobile sources (autos). However, environmental policy addresses many other issues, such as the use of pesticides and fertilizers by agriculture or deforestation). The problem of not in my back yard (NIMBY) is universal: we like the convenience of electricity but do not want its production or transport to affect our own property. Surface coalmines are an eyesore, but restoration can be costly. Hydroelectric dams can affect fisheries, flood unique canyons (causing a loss of scenic vistas), damage ecosystems (as in the Amazon), or displace human populations (as with China’s Three Gorges Project). Solar collection stations and wind generators require space and have impacts on aesthetics. For some, viewing large windmills along the crest of a lovely mountain range is an eyesore. For others, the same scene is a symbol of hope.

Environmental policy-makers must be aware of the relationship between changes in impacts in one medium and changes in impacts in other media. For example, reducing airborne emissions of mercury will also lead to reduced mercury concentrations in rivers and lakes. However, it may also be the case that reducing ozone precursors from auto emissions by using methyl tert-butyl ether (MTBE) leads to increasing harm to bodies of water as the MTBE precipitates out in rain. Finally, there may be policy and impact trade-offs that must be evaluated with reducing CO_2 emissions through a greater use of nuclear energy. The policy reduces greenhouse gases but raises issues and associated risks of waste storage and protection. In all cases, the links between different environmental media and different environmental policy must be understood for society to properly evaluate the trade-offs.

BURDEN OF PROOF

Because environmental issues tend to be complex, delays in responding to citizen concerns and new scientific information can lead to negative impacts or a local crisis. What is more problematic: erring on the side of environmental protection or erring on the side of development? When science is unclear or when studies yield conflicting outcomes, the issue of burden of proof arises. Two types of errors are possible. In a Type I error, a hypothesis is rejected when it is in fact true (e.g., deciding that a pollutant causes no health damages when in fact it does). Rejecting the hypothesis of a health link would lead to more emissions (and citizen exposure) than otherwise would be the case.

A Type II error occurs when the decision maker fails to reject a hypothesis that is in fact false (e.g., not rejecting the hypothesis that low doses of a pollutant have no damaging side effects for certain types of citizens, such as asthmatics, who are viewed as potentially sensitive to a particular pollutant). If in fact at low doses the pollutant does not have negative health impacts, environmental regulators might have imposed standards that induced costly compliance strategies that were based on the Type II error. Dose-response models that do not reject linear functions when the actual relationships are non-linear would fall into this category.

Both types of errors have costs. However, the political implications may depend on the type of error, leading decision-makers to prefer making errors that are difficult to detect. Thus, it can be argued that environmental regulators will tend to avoid making Type I errors. When evidence accumulates and shows conclusively that a pollutant has health impacts, those responsible for environmental policy do not want to be blamed for acting too slowly. Furthermore, citizens might prefer excessive caution (labeled a "precautionary bias"). On the other hand, Type II errors can result in regulators imposing high abatement costs onto polluters (and those purchasing associated products) in a manner that is not cost effective.

A related issue is whether or not the environmental impact is irreversible. If it is not reversible, a case can be made that the burden of proof should be assigned to those who assert that relatively higher levels of pollution are not problematic. On the other hand, if abatement costs are systematically underestimated and the benefits of pollution reduction are overestimated, it is possible to devote excessive resources to limiting environmental impacts.

ECONOMIC FRAMEWORK

Economists are aware that it is difficult to place monetary values on many impacts of pollution but argue that environmental amenities must be balanced against other valued goods and services.[3] Some view economists as over-emphasizing the efficacy of market incentives to the exclusion of other instruments. However, because economics offers a consistent framework for integrating insights from other fields, it will be described here.

Cost-Benefit Analysis (CBA)

The most fundamental economic analysis looks at how pollution impacts (reflected in "external costs") cause excessive consumption of polluting goods in the absence of government intervention. These external costs are the negative spillover effects of production or consumption for which no compensation is paid (e.g., a polluted stream that damages the health of those living along the stream). Producers consider the environment to be a free input; hence they only minimize private costs. If these external costs are added to the private costs (reflected in the supply curve), this is the total social cost.

Fig. 1 shows how a competitive product market yields an equilibrium price ($4) and quantity (80 units per week). However, in the absence of public intervention, the price only reflects the private costs of production, not damages imposed on others (amounting to $2 when the 80th unit is produced, but this is assumed to be less if only 65 units of the good are produced. The external costs are higher at higher levels of output presumably because damages rise dramatically when there are very high concentrations of the pollutant in the atmosphere). Determining the extent of those damages requires some valuation metric.

For now, let us assume that the analysts "got it right" in estimating both benefits and costs. This is a strong assumption because environmental services are notoriously hard to price. This problem can limit the ultimate effectiveness of CBA because the abatement costs tend to be short-term and quantifiable, but the benefits (avoided damages) are often long-term and difficult to quantify. For now, consider the impacts of environmental regulation within the CBA framework. Regulation requires pollution abatement activity, raising production costs but reducing the pollution and associated damages (as shown in Fig. 2).

The imposition of environmental regulation raises production costs (shifting the supply curve up) and reduces equilibrium consumption of the polluting good (from 80 to 75 per week) because the price has risen (from $4.00 to $4.40). In addition, external costs are reduced (so the sum of private and external costs is now $5 when 75 units of the good are produced). Emissions are reduced (though this particular figure only indicates the reduction in damages, not the precise reduction in emissions).

The next question is how much pollution abatement makes economic sense, since control costs rise rapidly as emissions are cut back towards zero. Continuing with our illustrative example, Default three depicts the total benefits of abatement and the total cost of abatement. The latter depends on the abatement technology and input prices and the interdependencies among production processes (for retrofitting control technologies). It is relatively easy to compute abatement costs from engineering cost studies, although predicting future control costs is not easy because innovations will create new control technologies. The benefits from abatement (or the

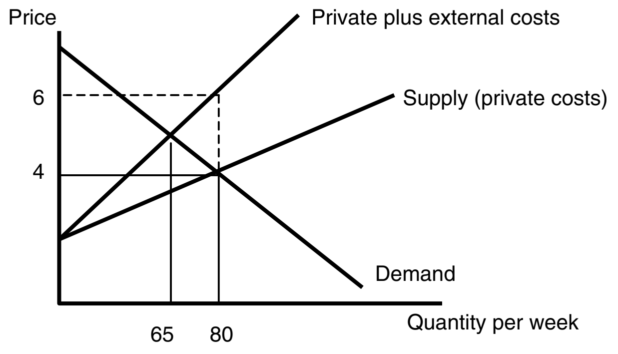


Fig. 1 Private costs and external costs.

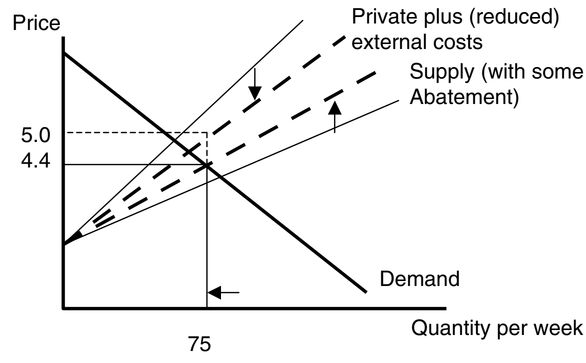


Fig. 2 Reducing external costs.

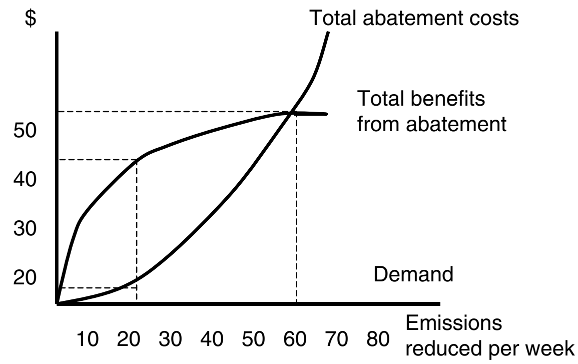


Fig. 3 Total benefits and total cost of abatement.

reduction of pollution damages—the cost of pollution) depend on the size of the affected population, incomes (indicating an ability to pay for environmental amenities), and citizen preferences (reflecting a willingness to pay). The benefits can be very difficult to estimate. Consider, for example, the health benefits of reduced particulates in the atmosphere, habitat values, and citizen valuations of maintaining a habitat for a particular species. Physical benefits can be found from dose-response studies. Various survey methods and market proxies for computing willingness to pay to avoid experiencing the impacts of pollution have methodological problems. However, if the dollar metric is to be used for determining the benefits of environmental improvements, techniques can at least establish rough benchmarks (as shown in Fig. 3) (Some argue strongly against the use of CBA.[4]).

The total benefits and costs to the marginal benefits and costs can be related because, for economists, the issue is not zero emissions versus unlimited emissions. In the former situation, if 80 units of the good results in 80 tn of emissions, then zero emissions reduced would be characterized as having no abatement costs (but also, no benefits from abatement). When the total benefits equal the total cost of abatement (at 65 tn of emissions reduced per week in Fig. 3), the last reductions in emissions were very costly and the additional benefits were fairly small. Zero

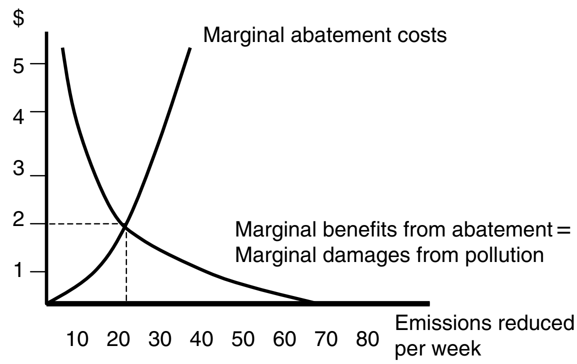


Fig. 4 Marginal benefits and marginal cost of abatement.

abatement activity would also be inefficient in this example because the marginal damages are very high and marginal abatement costs are quite low. Economics tries to determine the optimal amount of emissions. In the hypothetical example, the "optimal" quantity of reduced emissions is about 25, where the marginal abatement cost is just equal to the marginal benefits of $2. These are depicted in Fig. 4.

This outcome means that there are still 55 tn of emissions per week. If the estimated benefits and costs of pollution abatement are correct in this illustration, economic efficiency would be violated if additional resources were devoted to abatement activity. For example, if 50 tn of emissions were reduced (so only 30 tn of the pollutant are released), the marginal benefit would be about $1, but the marginal cost would be greater than $5. From the standpoint of economic efficiency, those resources could be used to create greater value in other activities.

Of course, the difficulty of obtaining a common dollar metric for all the impacts of different pollutants means that benefit-cost analysis must incorporate a range of values. The range could be incorporated in Fig. 4 as a band around the marginal benefit curve indicating one standard deviation from the calculated values. A conservative approach would recognize that the marginal benefit function could be above that depicted in Fig. 4, which would lead to optimal emission reduction of more than 25 tn per week (improving the ambient air quality).

Further complicating the analysis are production and exposure interdependencies. For example, the marginal cost of abatement associated with one type of emission may depend on the level of treatment (or abatement) for another contaminant. A joint optimization problem results, with the basic principles unchanged. Many investments in abatement equipment have this characteristic: once one set of contaminants is being reduced from a discharge flow and the cost of dealing with additional contaminants can be relatively low. For example, in the case of water discharges, if iron or manganese is removed via the precipitation method, total dissolved solids (TDS) is reduced and there may be an improvement in water clarity.

Interdependencies can also arise on the benefit side when the dose-response relationship for a particular contaminant is influenced by the presence of other contaminants. Again, in the case of secondary groundwater standards, perceptions of odor and color will be affected by whether or not they occur in combination. Such considerations must be factored into the analysis when comparing the benefits and costs of different treatment options.

Cost-Effective Analysis

Instead of trying to estimate the dollar benefits of saving a human life (or reducing the incidence of asthmatic attacks), one can compare the number of lives saved per dollar spent in abatement activity across programs. Thus, cost-effective analysis involves finding the least-cost method of achieving a given economic or social objective such as saving lives or retaining unique ecological settings. No dollar value (or explicit measure of avoided damages) is placed on that objective.[5] One advantage of this approach is that the focus is on minimizing the cost of meeting the (politically determined) target. It promotes consistency across a range of programs that might be designed to address a particular problem,whether that involves health impacts or a loss of habitat. Cost-effective analysis facilitates comparisons across programs, leading to reallocations of resources devoted to meeting such targets as new information is gathered over time.

POLICY INSTRUMENTS

Political systems have passed legislation and created agencies to apply laws to improve environmental performance. For example, in the United States, the Water Pollution Control Act of 1956 and the Clean Air Act of 1963 and subsequent amendments to both pieces of legislation have focused on achieving ambient standards. The U.S. Environmental Protection Agency is responsible for implementing these laws, and in other nations agencies have also been established to reduce emissions and improve environmental outcomes. A number of policy options can lead to emission reductions.[6,7] These instruments have different economic efficiency implications. In addition, some of these approaches are difficult to implement (due to being information-intensive), some are not cost effective (in that other approaches achieve the same outcome at lower cost), and the distributional implications can differ across these approaches (tax burdens differ or some groups obtain valuable assets).

Tax on the Polluting Good

An excise tax could be imposed on the good, cutting back consumption to 65 units per week (Fig. 1). Of course, the problem is not with the product but with the emissions associated with its production. Thus, this option does not provide incentives for developing new technologies that reduce abatement costs—it represents a static approach to the problem because it does not promote technological innovation.

Tax on Emissions

A penalty or charge for each ton of emissions would lead suppliers to cut back on emissions—to the extent that the abatement is less expensive than the tax. Thus, in Fig. 4, a tax of $2/tn would lead to the optimal reduction of pollutants. In addition, it provides incentives for innovation in the control technology industry. Firms will seek ways to reduce abatement costs, thus reducing their pollution taxes. This strategy is likely to be opposed by polluters who will be passing the taxes on to customers (where the ultimate incidence depends on supply and demand elasticities in the product market).

Tradable Emissions Permits

The same result (and incentive) is obtained if "allowances" of 25 tn are allocated to polluting firms, limiting emissions (the situation is not completely identical—a tax has certain costs to firms but yields uncertain overall abatement because regulators will not have precise estimates of abatement costs; the allowances have certainty in terms of overall abatement but uncertain cost. Of course, with monitoring, the tax can be varied over time to achieve a desired ambient condition). This approach provides an incentive for those with low abatement costs to reduce emissions and sell their permits (allowances) to others whose abatement costs would be very high. This places entrants at a disadvantage because incumbent firms are "given" these valuable allowances. The SO_2 regime in the United States has this feature. Of course, the initial allocations raise political issues (because permits represent wealth). In establishing a tradable permit regime, an environmental agency must determine the allowed level of emissions (here, 25 tn) and whether additional constraints might be applied to local areas with particular circumstances. In addition, the energy sector regulator has to make decisions regarding the treatment of cost savings from the regime. For example, savings might be passed on to consumers or retained by firms. The latter situation provides an incentive for utilities to participate in the emissions trade markets. A sharing plan can also be utilized so customers benefit as well.

Tighten Liability Rules

An alternative approach would utilize a court-based system, where fees would be assessed against those responsible for damaging the health of others, for reducing the economic value of assets, or for reducing the amenity values of ecosystems. Of course, this approach requires a well-specified set of property rights and clear causal links between specific emitters and affected parties. The transaction costs of such a system (resources devoted to negotiations and legal activity) could be prohibitive for many types of pollutants.

Emission Reduction Mandates (Quantity-Based Command-and-Control)

Although equal percentage cutbacks sound "fair," this strategy is not cost-effective because abatement costs can differ widely across pollution sources. If there are scale economies to emission reductions, it would be most efficient to have a few firms reduce emissions. The least-cost way to achieve a given overall reduction in emissions will involve differential cutbacks from different firms.

Mandate a Specific Control Technology (Technology-Based Command-and-Control)

This "command and control" strategy is not cost-effective because production conditions and retrofitting production processes differ across firms (based on the age of the plant and other factors). However, this policy option has been utilized in a number of situations as a "technology-forcing" strategy.

OTHER POLICY ISSUES

The above instruments have been utilized in different circumstances. Additional issues include intrinsic benefits, income distribution, sustainability, and renewable resources.

Intrinsic or Nonuse Benefits

Some people take a more expansive view of environmental amenities as they attempt to separate economic values from inherent values. However, this might be partly accounted for in terms of the perceived benefits to future generations. Intrinsic benefits from environmental programs include option values, existence values, and bequest values.[8] The first value represents a form of insurance so future access to a potential resource is not eliminated due to current consumption. The rationale behind option values is closely related to the "margin for error" argument noted earlier. Existence values reflect a willingness to pay for the knowledge that the amount of contaminant in the environment does not exceed particular levels or that a particular species (or level of biodiversity) is retained. The resource or ecological system is available for others. The bequest values can be interpreted as the willingness to pay for preserving a resource (or a geographic site) for future generations.

Redistributive Effects

It is important to note that citizens being harmed by emissions are not necessarily the same as those who are consuming the polluting good (such as electricity). Even if a particular program has positive net benefits, some parties are likely to be losers. They are seldom compensated and left better off, raising concerns about the distributional consequences of alternative policies. Furthermore, those harmed may have lower incomes (and thus, a lower willingness to pay to avoid damages due to the lower ability to pay). This point underscores the role of fairness as a factor that might outweigh efficiency considerations in some circumstances. Some agencies have been forbidden to use CBA on the grounds that the numbers are too speculative and that social concerns should be given priority. Intergenerational concerns can be interpreted as reflecting redistributive considerations.

Sustainable Development

Some of the issues associated with energy involve the use of nonrenewable resources (irreversibility). Some citizens argue that sustainability requires development that can be supported by the environment into the future. These people wish to ensure that resources are not depleted or permanently damaged. However, since sustainability depends on technology and innovations change resource constraints, defining the term with precision is quite difficult.

Renewable Energy Resources

Generating electricity without fossil fuels (e.g., hydro, wind, solar, biomass) is sometimes referred to as using green options. Green options are often limited in the amount (and reliability) of energy produced in a given time period. Utility applications for renewable resources include bulk electricity generation, on-site electricity generation, distributed electricity generation, and non-grid-connected generation. A number of regulatory commissions have required utilities to meet renewable portfolio standards. Such strategies reduce dependence on a particular energy source (to reduce the region's vulnerability to supply disruptions or rapid price run-ups). In addition, such requirements imply that managers are not making the most efficient investments in long-lived assets. Also, note that demand reduction through energy-efficient technologies is a substitute for energy, whatever the source.

CONCLUSIONS

The three main trends in environmental regulation in recent years have been shifting from command-and-control regulation towards a greater use of economic instruments (such as emissions trading), seeking more complete information on the monetary value of environmental costs and benefits, and a tendency for addressing environmental objectives in international meetings, as with the Kyoto Protocol.[9]

The interactions between economic and environmental regulation raise important policy issues. If energy sector regulation and environmental regulation remain separate, some means of harmonization may be necessary to promote improved performance. Collaboration would involve clarifying the economic duties of the environmental regulator and the environmental duties of the economic regulator. To avoid regulatory competition, agencies sometimes establish task forces or other mechanisms for identifying and resolving issues that might arise between jurisdictional boundaries (across states or between state and federal authorities). Such cooperation can serve to clarify the division of responsibilities and identify regulatory instruments that will most effectively meet economic and social objectives.

In summary, policy-makers respond to domestic political pressures by devising institutions and instruments to address pollution and environmental sustainability.[10] Although no single field of study contains all the tools necessary for sound policy formulation, economics does provide a comprehensive framework for evaluating the strengths and limitations of alternative policy options. Because of the pressures brought to bear by powerful stakeholders, adopted policies and mechanisms are not necessarily cost minimizing. The resulting inefficiencies may partly be due to considerations of fairness, which places constraints on whether, when, how, and where environmental impacts are addressed. As emphasized in this survey, citizens want to be good stewards of the land. We appreciate the adage: “The land was not given to us by our parents; it is on loan to us from our children.” How to be good stewards—through the development and implementation of sound environmental policies—has no simple answer given the complexity of the issues that need to be addressed.

REFERENCES

1. Maler, K.-G.; Vincent, J.R. In *Handbook of Environmental Economics: Environmental Degradation and Institutional Responses*, Vol. 1; North-Holland: Amsterdam, 2003.
2. Maler, K.-G.; Vincent, J.R. In *Handbook of Environmental Economics: Valuing Environmental Changes*, Vol. 2; North-Holland: Amsterdam, 2005.
3. Viscusi, W.K.; Vernon, J.M.; Harrington, J.E., Jr. *Economics of Regulation and Antitrust*; MIT Press: Cambridge, MA, 2000.
4. Ackerman, F.; Heinzerling, L. *Priceless: On Knowing the Price of Everything and the Value of Nothing*; New Press: New York, 2004.
5. Freeman, A.M., III *The Measurement of Environmental and Resource Values: Theory and Methods*, 2nd Ed.; Resources for the Future: Washington, DC, 2003.
6. Portney, P.; Stavins, R. *Public Policies for Environmental Protection*; Resources for the Future: Washington, DC, 2000.
7. Vig, N.; Kraft, M.E. *Environmental Policy and Politics: New Directions for the Twenty-first Century*, 5th Ed.; Congressional Quarterly Press: Washington, DC, 2003.
8. Krutilla, J. Conservation reconsidered. Am. Econ. Rev. **1967**, *57* (4), 777–786.
9. Busch, P.-O.; Jörgens, H.; Tews, K. The global diffusion of regulatory instruments: the making of a new international environmental regime In *The Annals of the American Academy of Political and Social Science*, Sage Publications: Thousand Oaks, CA, 2005; Vol. 598, 146–167.
10. Kraft, M.E. *Environmental Policy and Politics*, 3rd Ed.; Pearson/Longman: New York, 2004.

Evaporative Cooling

David Bisbee
Customer Advanced Technologies Program, Sacramento Municipal Utility District (SMUD), Sacramento, California, U.S.A.

Abstract
This report discusses the principles of evaporative cooling and provides an overview of the different types of systems currently on the market. It includes diagrams, system comparisons and a discussion of market barriers for direct evaporative coolers, indirect–direct evaporative coolers (IDECs) and the Coolerado Cooler.

INTRODUCTION

It is another sizzling summer day in America. Radio and television news weather forecasters wearily predict high temperatures of over 100°F all week long. Electric utilities all across the nation urge people to conserve electricity. Electricity demand and prices soar as air conditioners struggle to keep people comfortable. Although the energy crisis of 2001 may be over, the need for high efficiency air conditioning systems remains stronger than ever. This report discusses the principles of evaporative cooling and provides an overview of the different types of systems currently on the market.

PRINCIPLES OF EVAPORATIVE COOLING

When air comes into contact with water, some of the water evaporates. This happens because the temperature and the vapor pressure of the water and the air attempt to equalize. As the water molecules become a gas (evaporate), they "absorb" heat from the surrounding air and lower its temperature. The heat is still present, however, it has just been "captured" in the form of water vapor within the air (humidity). This phenomena is known as *adiabatic cooling* and is found throughout Nature—lakes, rivers, oceans, and anywhere there is water. Today people use evaporative cooling systems for a wide variety of purposes—comfort cooling, manufacturing processes, and other applications.

Some cool terms

Dry-bulb temperature: Temperature of air independent of its moisture content. It may be measured by using common thermometers.

Wet-bulb temperature: The lowest theoretical temperature achievable by evaporating water in a quantity of air at a given dry bulb temperature and humidity level, using only the heat within the air. It is measured by placing a moist piece of fabric over the thermometer and blowing air across it. The air evaporates the water, lowering the temperature on the thermometer to the wet bulb temperature[1].

Absolute humidity: The amount of water vapor present in air.

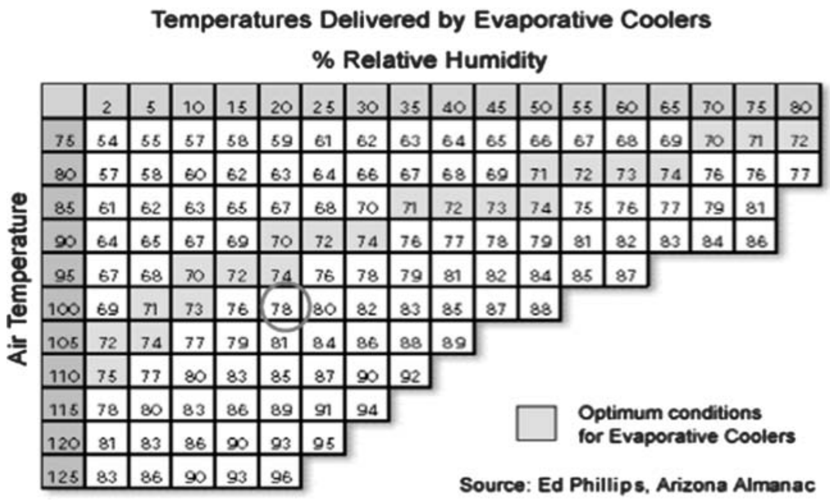
Relative humidity: The amount of water vapor present in air compared to the maximum amount it could hold.

Direct evaporative cooler: System in which the air comes into direct contact with wetted pads before entering the conditioned space. Commonly called swamp coolers.

Indirect-direct evaporative cooler (IDEC): System that uses a combination of direct evaporative cooling and an indirect evaporative heat exchanger. Commonly called two-stage evaporative coolers.

The effectiveness of evaporative cooling depends greatly upon the humidity or amount of water vapor in the air. Simply put, the drier the air, the more effective evaporative cooling will be. Air conditioning technicians often use a special thermometer known as a sling psychrometer to determine how much water vapor is present in the air. Essentially, this measurement involves placing a wet cloth (wick) on a thermometer and spinning the psychrometer until the water has evaporated and recording the lowest temperature obtained. This measurement is known as the "wet-bulb" temperature. What is important to understand about wet bulb temperature is that it tells us what the theoretical lowest supply air temperature for direct evaporative coolers will be. For example, standard design conditions for sizing residential

Keywords: Evaporative cooling; OASys; Coolerado Cooler; Swamp cooler; Direct evaporative coolers; Rigid pad coolers; Indirect–direct evaporative cooling; IDEC; IDAC; Maisotsenko cycle.



	2	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
75	54	55	57	58	59	61	62	63	64	65	66	67	68	69	70	71	72
80	57	58	60	62	63	64	66	67	68	69	71	72	73	74	76	76	77
85	61	62	63	65	67	68	70	71	72	73	74	75	76	77	79	81	
90	64	65	67	69	70	72	74	76	77	78	79	81	82	83	84	86	
95	67	68	70	72	74	76	78	79	81	82	84	85	87				
100	69	71	73	76	78	80	82	83	85	87	88						
105	72	74	77	79	81	84	86	88	89								
110	75	77	80	83	85	87	90	92									
115	78	80	83	86	89	91	94										
120	81	83	86	90	93	95											
125	83	86	90	93	96												

Fig. 1 Supply air temperatures delivered by direct evaporative coolers.
Source: From California Energy Commission (see Ref. 2).

cooling systems in Sacramento, California are:

- Outside air (a.k.a. "dry bulb") temperature: 100°F
- Wet bulb temperature: 70°F (relative humidity 23%)
- Desired indoor temperature: 78°F.

Under these conditions, the lowest theoretical supply air temperature that a standard evaporative cooler could provide would be 70°F. Realistically speaking, however, most direct evaporative cooling systems would only be able to achieve 78°F (Fig. 1). To put this into perspective, conventional refrigerant-based air conditioning systems are designed to deliver air at temperatures of about 55°F–65°F. At first glance, it appears that direct evaporative coolers would not work. After all, it is difficult to cool a home using 80°F air. To make matters worse, direct evaporative coolers increase the humidity within the house—making conditions even less comfortable. Fortunately, these shortcomings are somewhat mitigated by what is known as "effective temperature." Since evaporative coolers run at airflow rates that are 3–5 times higher than conventional air conditioning systems, they create a "cooling breeze" that makes the occupants of a room feel 4–6 degrees cooler than the actual temperature. For this reason, the effective temperature created by an evaporative cooler will feel 4–6 degrees cooler than the temperatures shown in the chart shown in Fig. 1. This chart also shows the dramatic effect that humidity has upon the performance of evaporative cooling systems. For this reason, direct evaporative cooling systems are limited to climates with low relative humidity levels. However, newer types of evaporative cooling systems, such as the OASys and the Coolerado Cooler may help somewhat overcome this limitation (more on this later).

Types of Evaporative Cooling Systems

There are three basic types of evaporative cooling systems: direct evaporative coolers (a.k.a. "swamp coolers"), indirect–direct evaporative coolers (IDECs), and the Coolerado Cooler.

Direct Evaporative Cooling Systems

In direct evaporative cooling systems, a.k.a. swamp coolers, the conditioned air comes into direct contact with the water used to produce the cooling effect. These systems are relatively simple and are widely used to provide comfort cooling for mobile homes, single-family housing, and industrial warehouses. Direct evaporative cooling systems generally cost about half as much to install as traditional vapor-compression systems and consume only about a fourth of the energy.[1] However, as discussed earlier, swamp coolers are limited to climates with dry air and produce high humidity levels within the conditioned space (Fig. 2). This is a considerable market

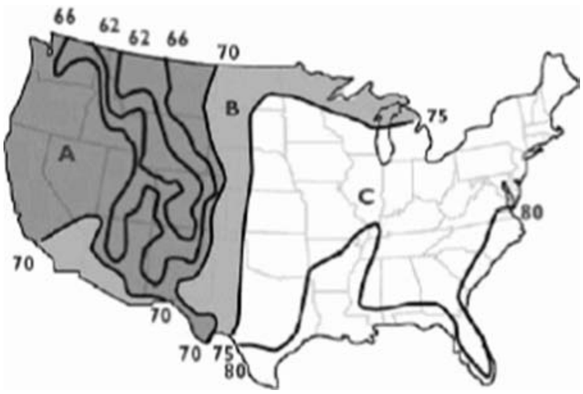


Fig. 2 Map showing climate zones suitable for direct evaporative coolers. Map is based upon wet bulb temperatures at 1% design conditions. Evaporative coolers would work best in the dry climates (areas marked A) may work somewhat in the areas marked B. However, in the eastern parts of the country, other types of air conditioners should be used.
Source: From U.S. Geological Survey (see Ref. 3).

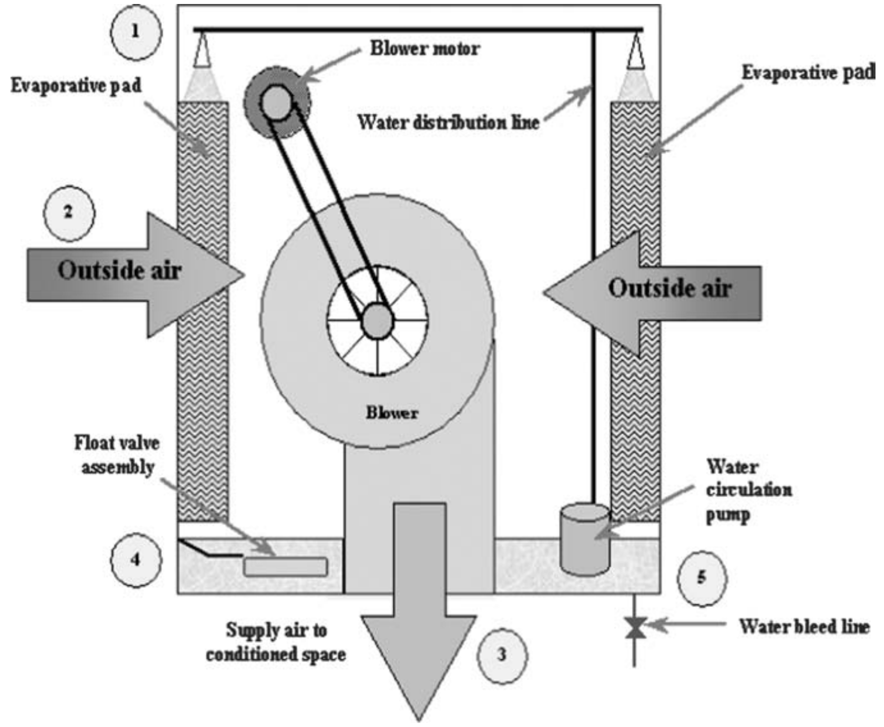


Fig. 3 Direct evaporative cooling system (fiber pad type).

barrier since studies conducted by American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) and others have consistently shown that most people prefer lower humidity levels.

Overview of Operation (Fig. 3)

1. Water is pumped from the basin (bottom of the unit) and distributed across the top of the evaporative cooling pads. These pads are usually made of wood shavings from aspen trees or other absorbent materials. Some of the water saturates the pads while the rest falls back down into the basin.
2. The blower draws outside air across the face of the evaporative cooling pads. When the air comes into contact with the wetted pads, some of the water evaporates and lowers the dry bulb temperature of the air.
3. The blower delivers cool moist air to the conditioned space. Since this air is very humid, it must ultimately be exhausted through open windows or relief dampers.
4. The evaporated water is replaced via a float valve assembly.
5. When water evaporates, minerals and other impurities are left behind. Eventually the water within the basin will become supersaturated with minerals and scale will begin to form on the surfaces of the cooler. To help address this issue, some water must be drained off (via the "bleed line") and replaced with fresh water. Modern systems accomplish this by periodically purging the basin water via a timer-controlled pump.

Not all Direct Evaporative Coolers are Created Equal

There two distinct types of direct evaporative coolers: fiber pad coolers and rigid sheet pad coolers.

1. *Fiber pad coolers*: (Fig. 4) the most commonly used evaporative cooling systems use pads that are made from shredded aspen wood fibers (a.k.a. excelsior) packed within a plastic net. Although synthetic fiber

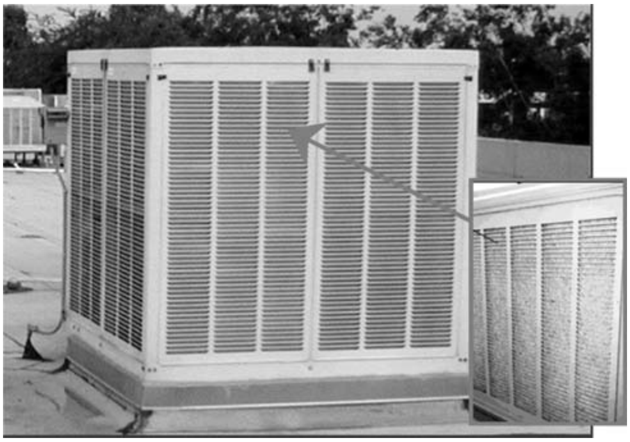


Fig. 4 Fiber pad evaporative cooler with aspen pads.

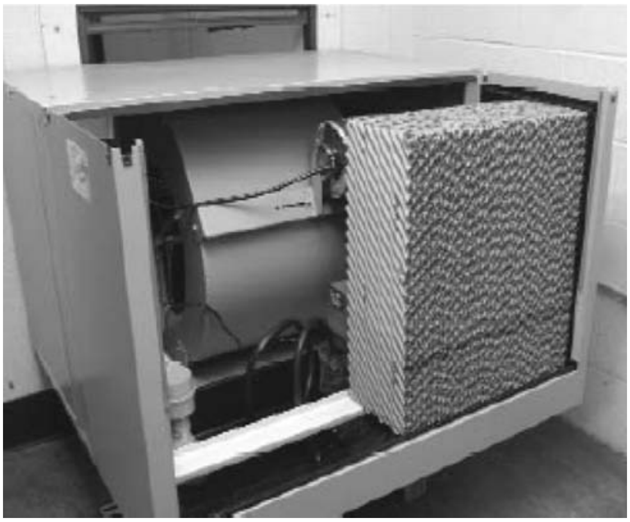


Fig. 5 Direct evaporative cooler with rigid sheet pads.
Source: From Pacific Gas & Electric Co. (see Ref. 5).

pads are available, they seldom perform as well as high-quality aspen pads. Fiber pads usually range from 1 to 2 in. in thickness. They are disposable and should be replaced every 1 or 2 years.[4]

2. *Rigid sheet pad coolers*: (Fig. 5) this type of cooler, also known as a "single-inlet cooler," used to be found only in large commercial applications, but is now becoming more common in residential applications as well. It uses rigid-sheet pads made from corrugated material that allows air to move through at higher velocities than is possible with aspen pads. Rigid sheet pads are usually 8 or 12 in. thick and have a corrugation pattern that forces water to flood the pad's air inlet side where most of the evaporation of water occurs. They can last for many years if water quality is properly maintained (Fig. 6).

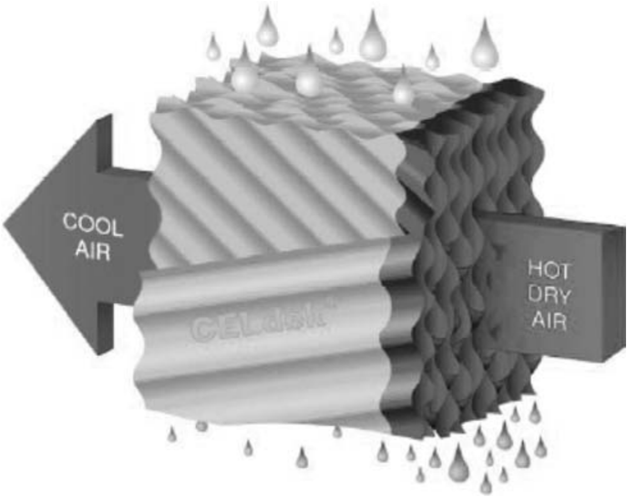


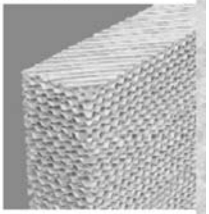
Fig. 6 Modern evaporative cooling media have to be efficient, which means that they must allow for as much cooling as temperature conditions allow while minimizing pressure drop, thereby saving fan power. Well-designed media filter the airstream but are also self cleaning in that the water dripping across the media to the sump below performs a cleaning function. Finally, the media should be durable and easy to replace at the end of their functional lifetime. Master Cool from Munters uses alternating corrugated layers to achieve these ends.
Source: From Munters Corporation (see Ref. 6).

Rigid sheet pad coolers are substantially more expensive than fiber pad coolers, but are much more energy efficient. According to a fact sheet published by Pacific Gas and Electric Company (PG&E), rigid sheet pad coolers can deliver air that is 5–10 degrees cooler than conventional aspen pad media coolers.[5] Both PG&E and Southern California Edison offered rebates to residential customers for qualifying "Energy Efficient Ducted Evaporative Coolers" (EEDEC) in 2005 (Fig. 7).

Finally, at least one prominent manufacturer, AdobeAir Inc., offers evaporative media (pads) that have been treated to resist microbial growth. This helps prevent the growth

Energy Efficient Ducted Evaporative Coolers (EEDEC)

EEDEC units are more efficient than conventional evaporative coolers because of the following advanced features:



Rigid Evaporative Media (typically 8-12" thick)

- Evaporative media which comes from manufacturers, and is certified with an 85% or higher evaporation efficiency delivers air that is 5 to 10 degrees cooler than the air delivered by a conventional aspen pad media evaporative cooler.
- Multi-function controls allow the EEDEC to also function as a whole house fan.
- Pressure relief dampers, if present, help in two ways; 1) they exhaust air to the outside typically through the attic keeping the attic cool which may reduce heat gain in the house, and 2) allow the system to operate when the house is not occupied.
- Special thermostatic controls remotely mounted to the wall and paired with pressure relief dampers allow for operation of the unit based on cooling requirements of the home.
- A water quality management pump eliminates old water from the system periodically and uses less water overall than continuous bleed systems.
- The EEDEC provides fresh air as it cools, using 100% outside air instead of re-circulating the existing indoor air as with conventional central air conditioning.

Fig. 7 Excerpt from a technical fact sheet produced by Pacific Gas & Electric Company (PG&E).
Source: From Pacific Gas & Electric Co. (see Ref. 5).

of microbes that can cause stains, odors, and product degradation.

Choosing the Right-sized Evaporative Cooler

Because direct evaporative coolers do not actually remove heat, they are rated by airflow instead of cooling capacity. According to PG&E, the following formulas may be used to determine the proper size cooler needed for residential applications[5]:

Hot dry climate: floor area (in square feet) × 4 CFM per square foot (based upon 30 air changes per hour and a ceiling height of 8 ft)

Average climate: floor area (in square feet) × 3 CFM per square foot (based upon 22.5 air changes per hour and a ceiling height of 8 ft)

Example: the size of an evaporative cooler needed for a 1500 square foot house in a hot dry climate would be: $1500\ ft^2 \times 4\ CFM/ft^2 = 6000$ cfm

Advantages of Direct Evaporative Cooling Systems

- Simple, proven technology: thousands of units have been sold and installed.
- Low first cost: generally cost less than half of conventional air conditioning systems.
- Energy efficient: 75% energy savings compared to conventional air conditioners.
- Easy to operate and service: units and replacement parts are widely available from multiple vendors, including many home improvement stores.
- Provides excellent ventilation: uses 100% outside air and may be used in a similar manner as a whole house fan (when operated in the fan only mode).

Disadvantages

- Manufacturers must overcome poor public perception caused by older, inferior equipment.
- Does not work well on humid days or in humid climates.
- Requires venting through open windows or relief dampers.
- Produces high humidity within conditioned space (most people prefer low humidity levels).
- Water quality has a major impact upon system performance and maintenance requirements. Hard water deposits (scale) may clog the media and reduce the airflow across the evaporative cooling pads (Fig. 8). Fortunately, new water quality management control strategies such as periodically purging the basin water have significantly reduced problems with scale.

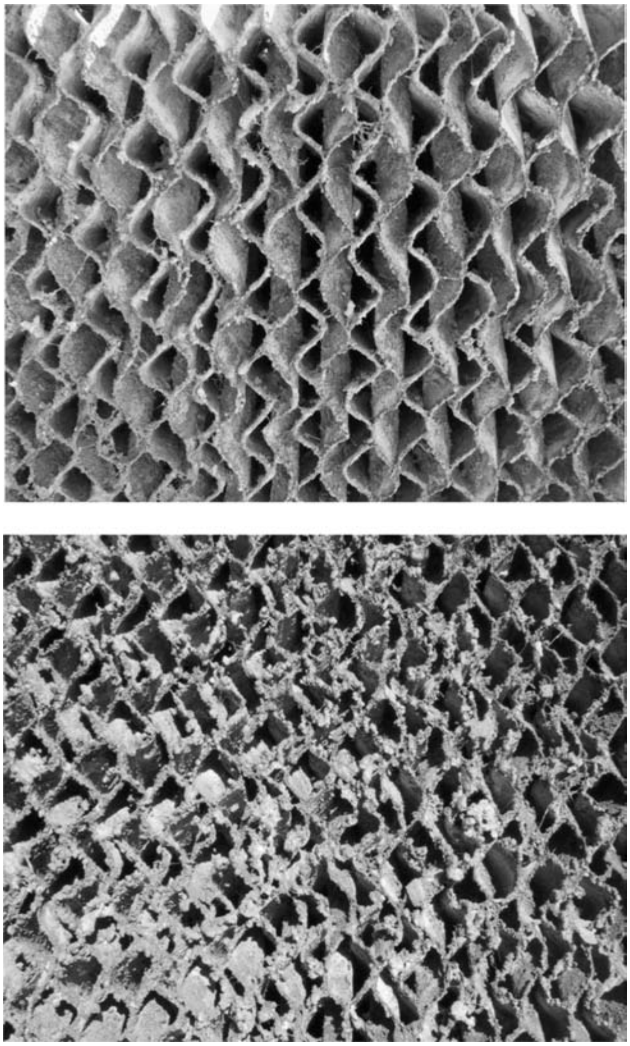


Fig. 8 Water quality has a major impact upon system performance and maintenance requirements. Note the hard-water deposits (scale) on the pads in the lower photo.

INDIRECT–DIRECT EVAPORATIVE COOLING SYSTEMS

As the name implies, IDEC or two-stage systems incorporate both direct evaporative cooling and an indirect evaporative heat exchanger. Unlike direct systems, the supply air within the heat exchanger does not come into direct contact with the water. Since this process reduces the amount of moisture added to the supply air, IDECs offer increased comfort over swamp coolers. Indirect–direct evaporative coolers cost about the same to install as conventional air conditioning systems, but are nearly four times more energy efficient.

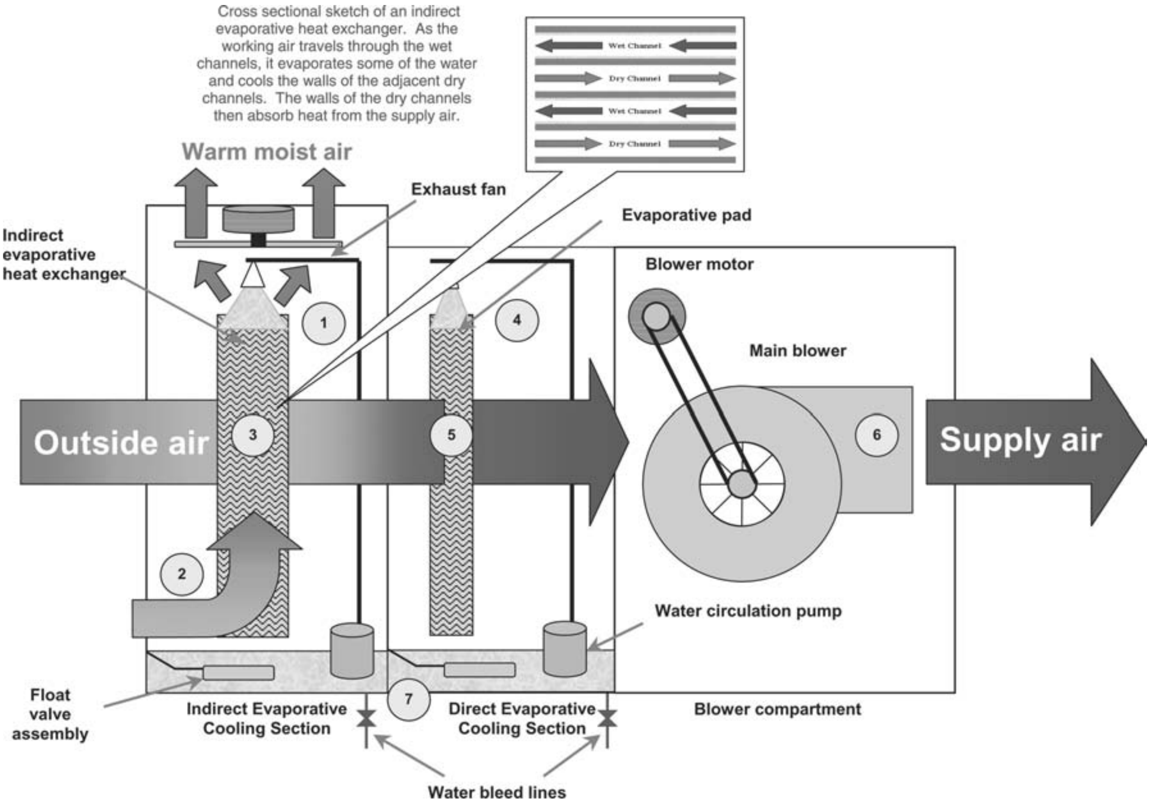


Fig. 9 Indirect–direct evaporative cooler (IDEC) system.

Overview of Operation (Fig. 9)

1. Water is pumped from the basin and distributed across the top of the indirect evaporative heat exchanger. The water clings to the surfaces within the wet channels of the heat exchanger as it travels back down into the basin.
2. The exhaust fan draws working air up through the wet channels. As the working air comes into contact with the water, some of the water evaporates and lowers the surface temperature of the heat exchanger. The warm moist working air is then exhausted through the top of the unit. The evaporated water is replaced via a float valve assembly.
3. The main blower draws outside air (supply air) across the face of the indirect evaporative heat exchanger. When the supply air comes into contact with the cool surface of the heat exchanger, some of its heat is transferred indirectly into the water via the heat exchanger. In other words, the supply air does not come into direct contact with the water. After the supply air passes through the indirect heat exchanger, it continues on to the direct evaporative cooling section.
4. Water is pumped from the basin and distributed across the top of the evaporative cooling pads. Some of the water saturates the pads while the rest falls back down into the basin of the unit.
5. The main blower draws the supply air across the face of the evaporative cooling pads. When the supply air comes into contact with the wetted pads, some of the water evaporates and lowers its dry bulb temperature.
6. Cool moist air is delivered into the conditioned space. Since this air is relatively humid, it must be exhausted through ducts or open windows.
7. The evaporated water is replaced via a float valve assembly. When the water evaporates, minerals and other impurities are left behind. Eventually the water within the basin will become super-saturated with minerals and scale will begin to form on the surfaces of the cooler. To help address this issue, some water must be drained off (via the bleed line) and replaced with fresh water. Modern systems accomplish this by periodically purging the basin water via a timer-controlled pump.

A Rough Beginning For IDECs

During the late 1990s, some early versions of IDECs were plagued with reliability and performance issues stemming

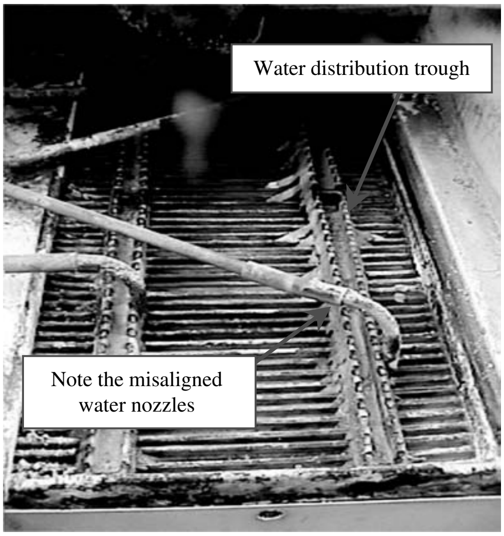


Fig. 10 The heat exchanger of this IDEC was rendered useless by misaligned water nozzles. Note also the presence of hard-water scale. This system was not equipped with a bleed line. Source: From Sacramento Municipal Utility District (see Ref. 7).

from a lack of understanding regarding proper system installation, misunderstood maintenance requirements and poorly designed systems.

Some examples:

1. *Reliability*: the water troughs of the indirect evaporative heat exchanger of the IDEC shown in Fig. 10 are designed to distribute water over the evaporative cooling media. Once the troughs are filled, the water is supposed to overflow across the top of the media. Unfortunately, this heat exchanger was rendered useless by misaligned water nozzles. This situation may have been easily avoided through the use of inexpensive hose clamps.
2. *Improper installation*: simply put, evaporative cooling systems require a lot of air—three to five times more than conventional vapor compression systems. Installers need to ensure that the ductwork is sized and installed correctly to ensure proper airflow is delivered to the conditioned space (Fig. 11).
3. *Poor engineering*: one particular IDEC system experienced severe reliablity problems due to faulty engineering. Periodically the water float valve would stick and cause the water within the basin to rise. Although there was an overflow drain to prevent the system from overflowing, the blower would draw water droplets into the motor and controls cabinet. Needless to say, water, motors and controls don't mix. Every time this occurred, it cost the owner about $500 to replace the motor and controls.

Fig. 11 A poorly installed IDEC system. The ductwork is far too restrictive to deliver the airflow needed to meet the cooling requirements of this classroom.
Source: From Sacramento Municipal Utility District (see Ref. 7).

Third Time a Charm?

The Davis Energy Group (DEG) and Speakman CRS have introduced a third generation IDEC known as the OASys that will hopefully overcome many of the problems of the past. The system was designed under a research grant provided by the California Energy Commission's Public Interest Energy Research (PIER) program. Additional tests were completed in 2005 in a field study sponsored by the Sacramento Municipal Utility District. According to the DEG, the OASys includes the following improvements (Fig. 12):

- Corrosion-proof, single-piece polyethylene cabinet ensures long life.
- Electronically commutated, variable-speed blower fan motor enhances efficiency and comfort during part-load conditions.
- Blower motor is located upstream of the evaporative cooling media—eliminates the chance of water droplets being drawn into the blower fan motor.
- Programmable thermostat that varies the speed of the blower motor in proportion to cooling load requirements.
- Water quality management controls (timed purge cycle).
- High performance evaporative cooling media and indirect heat exchanger enables OASys to cool

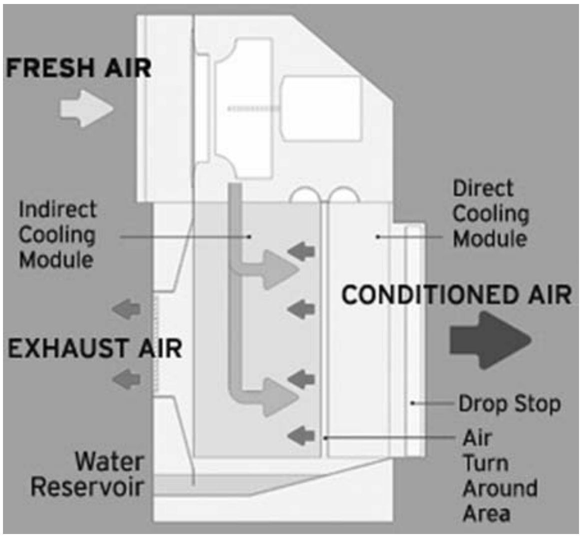


Fig. 12 Airflow diagram for the OASys.
Source: From Speakman CRS (see Ref. 10).

air *below* the ambient wet bulb temperatures (Fig. 13).

Advantages of IDECs

- Energy efficient—75% energy savings compared to conventional air conditioners.
- Offers increased comfort over swamp coolers (less moisture added to supply air).
- Provides excellent ventilation and may be used in a similar manner as a whole house fan (when operated in the fan only mode).

Disadvantages

- Does not work well in humid climates or on humid days.
- Units and replacement parts are not widely available.
- Requires venting through open windows or barometric relief dampers.
- Some previous IDECs have been known to have reliability problems.
- Adds some humidity to the conditioned space (most people prefer low humidity levels).
- Water quality has a major impact upon system performance and maintenance requirements. Hard water deposits (scale) may clog the media and reduce the airflow across the evaporative cooling pads. Fortunately, new water quality management control strategies such as periodically purging the basin water have significantly reduced problems with scale.

COOLERADO COOLER

A third type of evaporative cooling system was introduced in 2004—the Coolerado Cooler. This system utilizes a patented indirect evaporative heat exchanger that cools the air multiple times before it enters the

	Outdoor design conditions (°F)			Temperature of air delivered by unit (°F)		
	DB	WB	Product type	DB	WB	Resulting indoor RH (%) at 78°F
Sacramento, California	**100**	**70**				
			Coolerado	73.0	61.0	42.0
			OASys	67.0	64.7	60.8
			MasterCool + ICM	74.0	68.0	65.0
			Direct	75.1	70.0	70.5
Denver, Colorado	**93**	**60**				
			Coolerado	63.0	49.0	24.0
			OASys	56.7	52.8	37.3
			Direct	65.6	60.0	47.6

Notes: DB = drybulb temperature; F = Fahrenheit;
RH = relative humidity; WB = wetbulb temperature.
Assumptions for percent wetbulb achieved:
Coolerado Cooler = 90; OASys = 110;
MasterCool + ICM = 90; Direct = 83.

Source: E SOURCE; data from Davis Energy Group, Coolerado, and Pacific Gas and Electric [6]

Fig. 13 Cooling performance for different types of evaporative cooling systems. Note that the dry-bulb temperatures of the air delivered by the OASys are lower than the wet-bulb temperatures of the outside air.
Source: From E-Source (see Ref. 1).

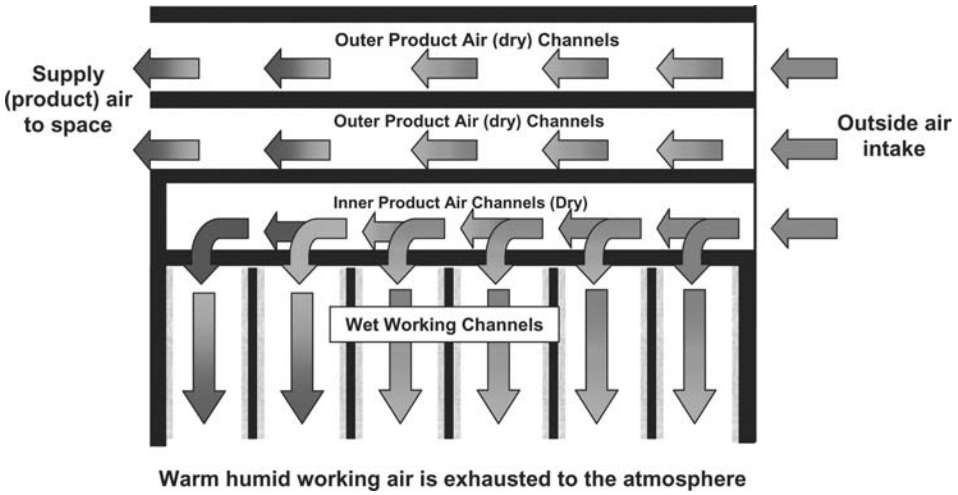


Fig. 14 The Maisotsenko cycle uses the same wet and dry channels as indirect evaporative coolers but with a much different geometry and airflow, creating a new thermodynamic cycle. It works by incrementally cooling and saturating working air and benefiting from that cooling on the next increment. This 2-D diagram of the Maisotsenko cycle shows how air is incrementally cooled by the continuous exhaust of heat followed by additional cooling. This cycle allows any fluid (gas or liquid) to be cooled below the wet-bulb temperature and within a few degrees of the dew-point temperature of the incoming working air. In addition, no moisture is added to the product airstream.
Source: From Idalex Technologies Inc. (see Ref. 8).

conditioned space. In essence this creates a "cascade" effect that enables the Coolerado Cooler to cool the air below the ambient air wet bulb temperature *without adding any moisture* to the supply air. The patented Coolerado heat and mass exchanger is designed to take advantage of the Maisotsenko Cycle (Fig. 14) and is constructed from cellulose fiber and ethyl vinyl acetate (EVA). The heat exchanger contains both wet and dry channels and separates incoming air into two air streams: Working Air and Product Air. The wet and dry air channels are separated via polyethylene coated sheets.

The Working Air is the air used to produce the cooling effect. As it travels through the wet channels of the heat exchanger, it comes into direct contact with wetted cellulose surfaces. When this happens, some of the water evaporates and absorbs heat (indirectly) from the Product Air stream (i.e., the air delivered to the conditioned space), which is traveling through the dry channel of the heat exchanger. This process is repeated several times within the heat exchanger.

Overview of Operation[9] (Fig 15)

1. Outside air is pushed into the Coolerado Cooler heat exchanger with a single fan.
2. Product air channels.
3. Working air channels.
4. Heat from the product air is transferred through the thin plastic and into the Wet Channels below.
5. Working air is blocked from entering the building.
6. The blocked working air is turned and passed through small holes into wet channels below the product air stream.

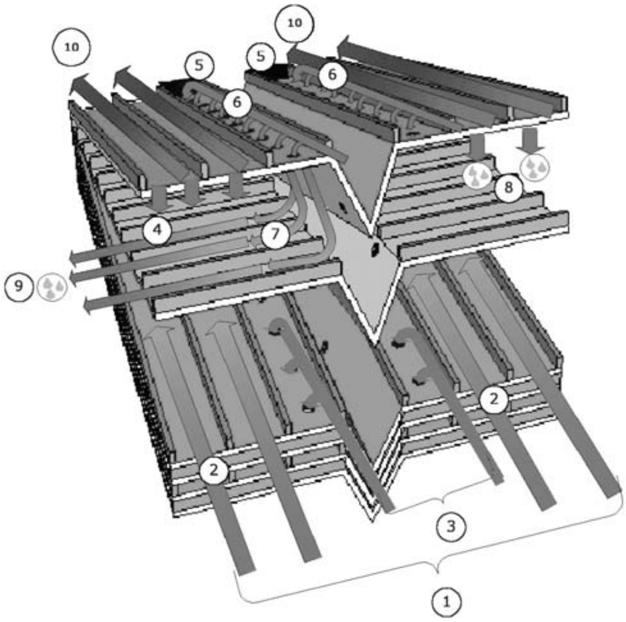


Fig. 15 Coolerado Cooler's patented heat-mass exchanger.
Source: From Idalex Technologies Inc. (see Ref. 9).

7. The working air is now moving through wet channels perpendicular or cross flow above and below the dry channels.
8. The heat that is passed from the dry channel is converted into water vapor.
9. Heat from the product air has been converted to water vapor and is now rejected as exhaust to the outside air.
10. The product air, which has now traveled the length of the heat exchanger, enters the conditioned space cold and dry.

Potential Applications

According to the manufacturer, the Coolerado Cooler may be used to provide residential and commercial comfort cooling or to pre-cool ventilation air for large conventional air conditioning systems in many areas of the United States. The map shown below is based upon the humidity within each climatic zone at the 1% design conditions (Fig. 16).

Advantages

- Energy efficient—75% energy savings compared to conventional air conditioners
- Does not add moisture to the conditioned air
- Able to cool near or below wet bulb temperature of the ambient air (Fig. 17).
- Provides excellent ventilation (Fig. 18).
- System has only two moving parts (supply fan and solenoid valve).

Disadvantages

- Units are not yet widely available.
- Requires venting through open windows or barometric relief dampers.

Fig. 16 Potential applications for the Coolerado Cooler in the continental United States.
Source: From Idalex Technologies Inc. (see Ref. 9).

	Outdoor design conditions (°F)			Temperature of air delivered by unit (°F)		
	DB	WB	Product type	DB	WB	Resulting indoor RH (%) at 78°F
Sacramento, California	**100**	**70**				
			Coolerado	73.0	61.0	42.0
			OASys	67.0	64.7	60.8
			MasterCool + ICM	74.0	68.0	65.0
			Direct	75.1	70.0	70.5
Denver, Colorado	**93**	**60**				
			Coolerado	63.0	49.0	24.0
			OASys	56.7	52.8	37.3
			Direct	65.6	60.0	47.6

Notes: DB = drybulb temperature; F = Fahrenheit;
RH = relative humidity; WB = wetbulb temperature.
Assumptions for percent wetbulb achieved:
Coolerado Cooler = 90; OASys = 110;
MasterCool + ICM = 90; Direct = 83.

Source: E SOURCE; data from Davis Energy Group, Coolerado, and Pacific Gas and Electric [6]

Fig. 17 Cooling performance for different types of evaporative cooling systems. Note that the dry-bulb temperatures of the air delivered by the coolerado are very close to the wet-bulb temperatures of the outside air.
Source: From E-Source (see Ref. 1).

- Reliability is unknown: fewer than 200 units have been deployed.
- Water quality may impact system performance and maintenance requirements.

SYSTEM COMPARISON

Recall that evaporative cooling systems incorporate adiabatic cooling. In this process, the heat within the air stream is converted into water vapor—it is not actually removed. Because of this, there is currently no established industry metric for comparing the performance of evaporative cooling systems to conventional vapor-compression systems. However, in April of 2005, E Source, a well-respected research firm, published a report on evaporative cooling systems. The chart shown in Fig. 19 provides an excellent summary of the relative performance and costs for all three types of evaporative cooling systems.

CONCLUSION

Evaporative cooling systems offer an energy efficient alternative to vapor-compression air conditioning systems in areas with favorable climatic conditions. Indeed as energy prices continue to climb, consumer demand for ultra-high efficiency air conditioning systems will increase. Innovative systems such as the OASys and the Coolerado Cooler may help overcome market barriers associated with the high humidity conditions provided by traditional swamp coolers. However, these systems will have to prove to be reliable and cost effective in order to compete with well-established refrigerant-based air conditioning systems.

Fig. 18 Coolerado Cooler installation at a school in Sacramento, California.
Source: From Sacramento Municipal Utility District (see Ref. 7).

Manufacturer	Model	Type	Adds moisture to the air?	WB achievable (%)[a]	EER[b]	Approximate price ($)	Peak draw (kW)
Evaporative coolers							
Coolerado	Coolerado Cooler R600	Indirect	No	90–95[c]	40–65[c]	3,000 wholesale	1.30
Speakman CRS	OASys	IDEC	Yes	109–116[c]	40–136[c]	2,500 retail	0.63
Adobe Air	Mastercool + ICM	IDEC	Yes	88–98[c]	xx	1,473 retail	0.47
Phoenix Manufacturing Inc.	PD 4801/PD 4231	Direct	Yes	Up to 93.5[d]	xx	700–900 wholesale	xx
Air conditioners							
Goodman	CLJ36-1 ARUF42	DX	No	NA	11	1,000–2,000 wholesale	3.30
Goodman	CLQ36 AEPT36	DX	No	NA	13.8	1,500–2,500 wholesale	2.60

Notes: cfm = cubic feet per minute; DX = direct expansion; EER = energy-efficiency ratio; CM = Indirect Cooling Module; IDEC = indirect/direct evaporative cooler; kW = kilowatt; NA = not applicable; SEER = seasonal energy-efficiency ratio; WB = wetbulb temperature; xx = not available.

a. Wetbulb achievable percentage is the percentage of the difference between the drybulb and wetbulb temperature by which the unit can lower the drybulb temperature.

b. EERs listed here for evaporative coolers will vary with airflow and climate and are applicable primarily in the western United States. For the OASys, this number is a sensible EER and is not directly comparable with the Coolerado Cooler efficiency.

c. Based on testing performed by National Renewable Energy Laboratory (Coolerado Cooler) and Davis Energy Group; certified by Lawrence Berkeley National Laboratory (OASys), Pacific Gas and Electric (Mastercool + ICM).

d. Manufacturer's claim.

Source: E SOURCE; based on manufacturers' data and test data.

Fig. 19 Comparison of different types of evaporative cooling systems.
Source: From E-Source (see Ref. 1).

REFERENCES

1. Criscione, P. *Death of the Swamp Thing: Evaporative Cooling Delivers Dry Air*, ET Currents Number 41, April 2005; 1–6. E Source, 1965 North 57th Court, Boulder, CO 80301. USA, Tel 303.444.7788 x109, www.esource.com.
2. Evaporative Cooling. California Energy Commission. Available at: http://www.consumerenergycenter.org/home/heating_cooling/evaporative.html (accessed on December 6, 2006).
3. U.S. Geological Survey. Available at: http://ga.water.usgs.gov/edu/watercycleevaporation.html (accessed on December 6, 2006).
4. Installing and Maintaining Evaporative Coolers, by Roy otterbein, Home Energy Magazine Online, May/June 1996. Available via archives at: http://www.homeenergy.org/index.php (accessed on December 6, 2006).
5. Pacific Gas & Electric Co. Energy-Efficient Ducted Evaporative Coolers, Available at: http://www.pge.com/docs/pdfs/res/rebates/EEDECTechSheetv5.pdf (accessed on December 6, 2006).
6. Graphic provided courtesy of Munters Corporation, Humicool Sales & Manufacturing, 108 Sixth Street, Fort Myers, FL 33907, Tel: (239) 936-1555.
7. Bisbee, D. *Operation IDEC Rescue*. Sacramento Municipal Utility District, 6301 S Street, Sacramento, CA 95817-1899, Tel (916) 732-6409, available at: http://www.smud.org/education/cat/index.html (accessed on December 6, 2006).
8. Idalex Technologies, Inc., 4700 W 60th Ave., Unit 3, Arvada, CO USA 80003 tel (303) 375-0878 available at: http://www.idalex.com/technology/how_it_works_-_engineering_perspective.htm (accessed on December 6, 2006).
9. How the Coolerado Cooler Works. Available at: http://www.coolerado.com (accessed on December 6, 2006), Coolerado LLC, 4700 West 60th Avenue, Unit 3, Arvada, CO, 80003, Tel: 303-375-0878.
10. FAQ's Speakman CRS, 400 Anchor Mill Road, New Castle, DE 19899-0191, U.S.A., Tel: (302) 351-2416. Available at: http://www.oasysairconditioner.com/faq.html (accessed on December 6, 2006).

Exergy: Analysis

Marc A. Rosen
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Exer—Fed

Abstract
Exergy analysis is an assessment technique for systems and processes that is based on the second law of thermodynamics. Exergy analysis has been increasingly applied over the last several decades largely because of its advantages over energy analysis: more meaningful efficiencies are evaluated because exergy efficiencies are always a measure of the approach to the ideal and inefficiencies in a process are better pinpointed because exergy analysis quantifies the types, causes, and locations of losses. In this article, the role of exergy analysis in assessing and improving energy systems is examined. Also, exergy and its use as an analysis technique are briefly described, the range of energy systems that have been assessed with exergy analysis are surveyed, and several example applications of exergy analysis are presented, ranging from simple to complex.

INTRODUCTION

Energy analysis is based on the first law of thermodynamics, which embodies the principle of conservation of energy and is the traditional method used to assess the performance and efficiency of energy systems and processes.

Exergy analysis is a thermodynamic analysis technique for systems and processes that is based on the second law of thermodynamics. Exergy analysis has been increasingly applied over the last several decades, in large part because of its advantages over energy analysis:

- More meaningful efficiencies are evaluated with exergy analysis because exergy efficiencies are always a measure of the approach to the ideal.
- Inefficiencies in a process are better pinpointed with exergy analysis because the types, causes, and locations of the losses are identified and quantified.

In this article, the role of exergy analysis in the assessment and improvement of energy systems is examined. First, exergy and its use as an analysis technique are briefly described. Second, the ranges of energy systems that have been assessed with exergy analysis are surveyed. Third, several example applications of exergy analysis are presented, ranging from simple devices to large and complex systems.

Keywords: Exergy; Irreversibility; Efficiency; Second law of thermodynamics; Entropy; Environment; Economics; Design.

EXERGY

Exergy can be regarded as a measure of the usefulness or quality of energy. Technically, exergy is defined as the maximum amount of work that can be produced by a stream of energy or matter, or from a system, as it is brought into equilibrium with a reference environment. Unlike energy, exergy is consumed during real processes due to irreversibilities and conserved during ideal processes. Exergy and related concepts have been recognized for more than a century.[1]

Exergy analysis is a methodology that uses the first and second laws of thermodynamics for the analysis, design, and improvement of energy and other systems.[2–14] The exergy method is useful for improving the efficiency of energy-resource use, for it quantifies the locations, types, and magnitudes of wastes and losses. In general, more meaningful efficiencies are evaluated with exergy analysis rather than energy analysis because exergy efficiencies are always a measure of the approach to the ideal. Therefore, exergy analysis accurately identifies the margin available to design more efficient energy systems by reducing inefficiencies.

In evaluating exergy, the characteristics of the reference environment must be specified,[2–15] usually by specifying the temperature, pressure, and chemical composition of the reference environment. The results of exergy analyses, consequently, are relative to the specified reference environment, which in most applications is modeled after the actual local environment. The exergy of a system is zero when it is in equilibrium with the reference environment. The tie between exergy and the environment has implications regarding environmental impact.[7,8,16]

The theory and the applications of exergy have been described in specialized books, e.g.,[2–8] general thermodynamics texts, e.g.,[9,10] and journal articles e.g.[11–14] Many applications of exergy analysis have been reported in fields ranging from power generation[17,18] and cogeneration[9] to district energy,[19] thermal processes,[20,21] and thermal energy storage[22,23] and on to systems as large as countries.[24,25]

Exergy and the Reference Environment

Exergy quantities are evaluated with respect to a reference environment. The intensive properties of the reference environment in part determine the exergy of a stream or system. The reference environment is in stable equilibrium, with all parts at rest relative to one another and with no chemical reactions occurring between the environmental components. The reference environment acts as an infinite system and is a sink and source for heat and materials. It experiences only internally reversible processes in which its intensive state remains unaltered (i.e., its temperature T_0, pressure P_0, and the chemical potentials μ_{i00} for each of the i components present remain constant). The exergy of the reference environment is zero. More information on reference-environment models can be found in this encyclopedia in an article by the present author entitled "Exergy: Environmental Impact Assessment Applications."

Exergy Balances

Energy and exergy balances can be written for a general process or system.

Since energy is conserved, an energy balance for a system may be written as

$$\text{Energy input} - \text{Energy output} = \text{Energy accumulation} \quad (1)$$

Energy input and output refer respectively to energy entering and exiting through system boundaries. Energy accumulation refers to build-up (either positive or negative) of the quantity within the system.

By contrast, an exergy balance can written as

$$\text{Exergy input} - \text{Exergy output} - \text{Exergy consumption} = \text{Exergy accumulation} \quad (2)$$

This expression can be obtained by combining the principles of energy conservation and entropy nonconservation, the latter of which states that entropy is created during a process due to irreversibilities. Exergy is consumed due to irreversibilities, with exergy consumption proportional to entropy creation.

Eqs. 1 and 2 demonstrate an important difference between energy and exergy—energy is conserved while exergy, a measure of energy quality or work potential, can be consumed.

Definitions

It is helpful to define some terms related to exergy for readers. The following are exergy quantities:

Exergy: A general term for the maximum work potential of a system, a stream of matter, or a heat interaction in relation to the reference environment (see definition below) as the datum state; or the maximum amount of shaft work obtainable when a steady stream of matter is brought from its initial state to the dead state (see definition below) by means of processes involving interactions only with the reference environment.

Physical exergy: The maximum amount of shaft work obtainable from a substance when it is brought from its initial state to the environmental state (see definition below) by means of physical processes involving interaction only with the environment.

Chemical exergy: The maximum work obtainable from a substance when it is brought from the environmental state to the dead state by means of processes involving interaction only with the environment.

Thermal exergy: The maximum amount of shaft work obtainable from a given heat interaction using the environment as a thermal energy reservoir.

Exergy consumption: The exergy consumed during a process due to irreversibilities within the system boundaries.

The following terms relate to the reference environment and its state:

Reference environment: An idealization of the natural environment, which is characterized by a perfect state of equilibrium, i.e., absence of any gradients or differences involving pressure, temperature, chemical potential, kinetic energy, and potential energy. The reference environment constitutes a natural reference medium with respect to which the exergy of different systems is evaluated.

Dead state: The state of a system when it is in thermal, mechanical, and chemical equilibrium with a conceptual reference environment, which is characterized by a fixed pressure, temperature, and chemical potential for each of the reference substances in their respective dead states.

Environmental state: The state of a system when it is in thermal and mechanical equilibrium with the reference environment, i.e., at the pressure and temperature of the reference environment.

Reference state: A state with respect to which values of exergy are evaluated. Several reference states are used, including environmental state, dead state, standard environmental state, and standard dead state.

EXERGY ANALYSIS

Exergy analysis involves the application of exergy concepts, balances, and efficiencies to evaluate and improve energy and other systems. Many engineers and scientists suggest that devices are best evaluated and improved upon using exergy analysis in addition to or in place of energy analysis.

A journal devoted to exergy matters entitled *The International Journal of Exergy* has recently been established by Inderscience. Some extensive bibliographies have been compiled, including one by Goran Wall (see the Web site http://exergy.se).

A simple procedure for performing energy and exergy analyses involves the following steps:

- Subdivide the process under consideration into as many sections as desired, depending on the depth of detail and the understanding desired from the analysis.
- Perform conventional mass and energy balances on the process and determine all basic quantities (e.g., work, heat) and properties (e.g., temperature, pressure).
- Based on the nature of the process, the acceptable degree of analysis complexity and accuracy, and the questions for which answers are sought, select a reference-environment model.
- Evaluate energy and exergy values relative to the selected reference-environment model.
- Perform exergy balances, including the determination of exergy consumptions.
- Select efficiency definitions depending on the measures of merit desired and evaluate the efficiencies.
- Interpret the results and draw appropriate conclusions and recommendations relating to such issues as design changes, retrofit plant modifications, etc.

EXERGY ANALYSIS AND EFFICIENCY

Increases in efficiency are subject to two constraints, which are often poorly understood:

- Theoretical limitations, which establish the maximum efficiency theoretically attainable for a process by virtue of the laws of thermodynamics
- Practical limitations, which further limit increases in efficiency

First, consider practical limitations on efficiency. In practice, the goal when selecting energy sources and utilization processes is not to achieve maximum efficiency, but rather to achieve an optimal trade-off between efficiency and such factors as economics, sustainability, environmental impact, safety, and societal and political acceptability. This optimum is dependent on many factors controllable by society. Furthermore, these factors can be altered to favor increased efficiency (e.g., governments can offer financial incentives that render high-efficiency technologies economically attractive or provide disincentives for low-efficiency alternatives through special taxes and regulations).

Next, consider theoretical limitations on efficiency, which must be clearly understood to assess the potential for increased efficiency. Lack of clarity on this issue in the past has often led to confusion, in part because energy efficiencies generally are not measures of how nearly the performance of a process or device approaches the theoretical ideal. The consequences of such confusion can be significant. For example, extensive resources have at times been directed towards increasing the energy efficiencies of devices that in reality were efficient and had little potential for improvement. Conversely, devices at other times have not been targeted for improved efficiency even though the difference between the actual and maximum theoretical efficiencies, which represents the potential for improvement, has been large.

The difficulties inherent in energy analysis are also attributed to the fact that it only considers quantities of energy and ignores energy quality, which is continually degraded during real processes. Exergy analysis overcomes many of the problems associated with energy analysis.

OVERVIEW OF EXERGY ANALYSIS APPLICATIONS

Exergy analysis has been applied to a wide range of processes and systems, including those that are mechanical, thermal, electrical, and chemical. The types of applications of exergy methods that have been reported over the last several decades include:

- Electricity generation using both conventional devices such as fossil and nuclear power plants as well as alternative devices such as fuel cells and solar energy systems.
- Energy storage systems such as batteries, pumped storages, and thermal energy storages.
- Combustion technologies and systems and engines of various types.
- Transportation systems for land, air, and water transport.
- Heating and cooling systems for building systems and industrial applications.
- Cogeneration systems for producing heating and electrical needs simultaneously.
- Chemical processes such as sulphuric acid production, distillation, and water desalination, as well as petrochemical processing and synthetic fuels production.
- Metallurgical processes such as lead smelting.

EXAMPLES OF EXERGY ANALYSIS APPLICATIONS

Three examples of differing complexity of applications of exergy analysis are presented:

- An electrical resistance space heater (a simple component)
- A thermal energy storage system (a simple system containing a number of components)
- A coal-fired electrical generating station (a complex system)

Electrical Resistance Space Heater

An electrical resistance space heater converts electricity to heat at a temperature suitable for room comfort, and is illustrated in Fig. 1 (Part A).

The energy efficiency of electric resistance space heating often exceeds 99%, implying that the maximum possible energy efficiency for electric resistance heating is 100%, corresponding to the most efficient device possible.

This understanding is erroneous; however, energy analysis ignores the fact that in this process, high-quality energy (electricity) is used to produce a relatively low-quality product (warm air). Exergy analysis recognizes this difference in energy qualities and indicates the exergy of the heat delivered to the room to be about 5% of the exergy entering the heater. Thus, the exergy efficiency of electric resistance space heating is about 5%.

The exergy results are useful. Since thermodynamically ideal space heating has an exergy efficiency of 100%, the same space heating can in theory be achieved using as little as 5% of the electricity used in conventional electric resistance space heating. In practical terms, one can achieve space heating with a greatly reduced electricity input using an electric heat pump (see Part B of Fig. 1), using 15% of the electricity that electric resistance heating would require, for a heat pump with a "coefficient of performance" of 7.

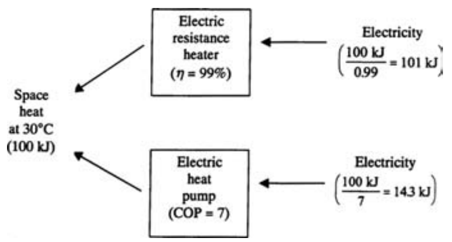


Fig. 1 Comparison of the quantity of electricity required to provide 100 kJ of space heat using two different heating devices: (a) an electric resistance heater and (b) an electric heat pump. Here, η denotes energy efficiency and COP coefficient of performance.

Thermal Storage System

A thermal energy storage system receives thermal energy and holds it until it is required. Thermal storages can store energy at temperatures above or below the environment temperature, and they come in many types (e.g., tanks, aquifers, ponds, caverns).

The evaluation of a thermal energy storage system requires a measure of performance which is rational, meaningful, and practical. The conventional energy storage efficiency is inadequate. A more perceptive basis is needed if the true usefulness of thermal storages is to be assessed and their economic benefit optimized, and exergy efficiencies provide such performance measures.

The notion that energy efficiency is an inappropriate measure of thermal storage performance can be illustrated. Consider a perfectly insulated thermal storage containing 1000 kg of water, initially at 40°C. The ambient temperature is 20°C and the specific heat of water is taken to be constant at 4.2 kJ/kg K. A quantity of 4200 kJ of heat is transferred to the storage through a heat exchanger from an external body of 100 kg of water cooling from 100 to 90°C. This heat addition raises the storage temperature 1.0°C to a value of 41°C. After a period of storage, 4200 kJ of heat are recovered from the storage through a heat exchanger, which delivers it to an external body of 100 kg of water, raising the temperature of that water from 20 to 30°C. The storage is returned to its initial state at 40°C.

For this storage cycle, the energy efficiency—the ratio of heat recovered from the storage to heat injected—is 4200/4200 kJ = 1, or 100%. But the recovered heat is at only 30°C and is of little use, having been degraded even though the storage energy efficiency was 100%. The exergy recovered in this example is 70 kJ and the exergy supplied 856 kJ. Thus the exergy efficiency, the ratio of the thermal exergy recovered from storage to that injected, is 70/856 = 0.082 or 8.2%, a much more meaningful expression of the achieved performance.

Consequently, a device which appears to be ideal on an energy basis is correctly shown to be far from ideal on an exergy basis, clearly demonstrating the benefits of using exergy analysis for evaluating thermal storage.

Coal-Fired Electrical Generating Station

Energy and exergy analyses are applied to the Nanticoke coal-fired electrical generating station in Ontario, Canada, which has a net unit electrical output of approximately 500 MWe and is operated by the provincial electrical utility, Ontario Power Generation (formerly Ontario Hydro). This example illustrates how exergy analysis allows process inefficiencies to be better pinpointed than an energy analysis does and how efficiencies are to be more rationally evaluated.

A detailed flow diagram for a single unit of the station is shown in Fig. 2. The symbols identifying the streams are described in Table 1a–c for material, thermal, and electrical flows, respectively, with corresponding data. Fig. 2 has four main sections:

- *Steam Generation*. Eight pulverized-coal-fired natural circulation steam generators each produce 453.6 kg/s steam at 16.89 MPa and 538°C and 411.3 kg/s of reheat steam at 4.00 MPa and 538°C. Air is supplied to the furnace by two 1080 kW 600-rpm motor-driven forced draft fans. Regenerative air preheaters are used. The flue gas passes through an electrostatic precipitator rated at 99.5% collection efficiency and exits the plant through two multiflued, 198 m high chimneys.
- *Power Production*: The steam passes through a series of turbine generators linked to a transformer. Extraction steam from several points on the turbines preheats feedwater in several low- and high-pressure heat exchangers and one spray-type open deaerating heat exchanger. The low-pressure turbines exhaust to the condenser at 5 kPa. Each station unit has a 3600-rpm, tandem-compound, impulse-reaction turbine generator containing one single-flow high-pressure cylinder, one double-flow intermediate-pressure cylinder, and two double-flow low-pressure cylinders. Steam exhausted from the high-pressure cylinder is reheated in the combustor.
- *Condensation*: Cooling water from Lake Erie condenses the steam exhausted from the turbines. The cooling-water flow rate is adjusted to achieve a specified cooling-water temperature rise across the condenser.
- *Preheating*: The temperature and pressure of the feedwater are increased in a series of pumps and feedwater-heater heat exchangers.

The reference-environment model used here has a temperature of 15°C (the approximate mean temperature of the lake cooling water), a pressure of 1 atm, and a chemical composition consisting of air saturated with water vapor, and the following condensed phases at 15°C and 1 atm: water (H_2O), gypsum ($CaSO_4 \cdot 2H_2O$), and limestone ($CaCO_3$). For simplicity, heat losses from external surfaces are assumed to occur at the reference-environment temperature of 15°C.

Energy and exergy values for the streams identified in Fig. 2 are summarized in Table 1. Exergy-consumption values for the devices are listed, according to process

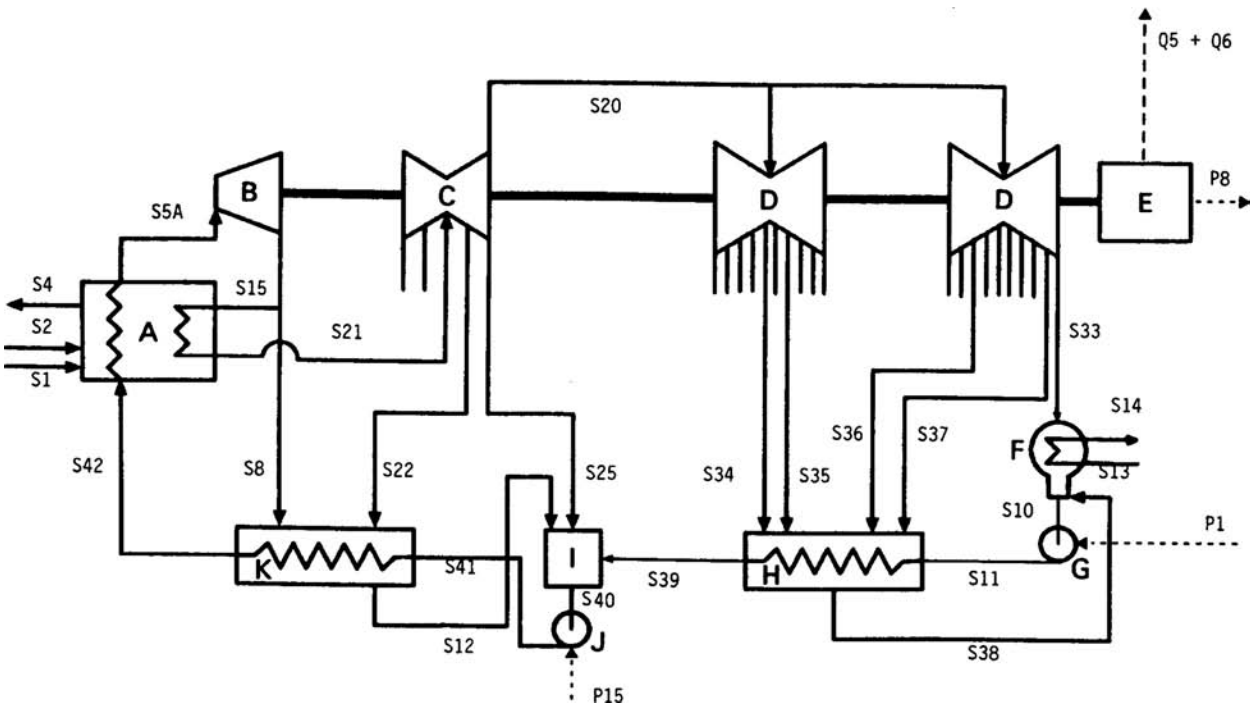


Fig. 2 A unit of the coal-fired electrical generating station. Lines exiting the turbines represent extraction steam. The station has four main sections: *Steam Generation* (Device A), *Power Production* (B–E), *Condensation* (F), and *Preheating* (G–K). The external inputs for Device A are coal and air, and the output is stack gas and solid waste. The external outputs for Device E are electricity and waste heat. Electricity is input to Device G and Device J and cooling water enters and exits Device F. A: steam generator and reheater, B: high-pressure turbine, C: intermediate-pressure turbine, D: low-pressure turbines, E: generator and transformer, F: condenser, G: hot well pump, H: low-pressure heat exchangers, I: open deaerating heat exchanger, J: boiler feed pump, K: high-pressure heat exchangers.

Exer—Fed

Table 1a Data for material flows for a unit of the coal-fired electrical generating station

Stream	Mass flow rate (kg/s)[a]	Temperature (°C)	Pressure (N/m^2)	Vapor frac.[b]	Energy flow rate (MW)	Exergy flow rate (MW)
S1	41.74	15.00	1.01×10^5	solid	1367.58	1426.73
S2	668.41	15.00	1.01×10^5	1.0	0.00	0.00
S3[c]	710.15	1673.59	1.01×10^5	1.0	1368.00	982.85
S4	710.15	119.44	1.01×10^5	1.0	74.39	62.27
S5A	453.59	538.00	1.62×10^7	1.0	1585.28	718.74
S8	42.84	323.36	3.65×10^6	1.0	135.44	51.81
S10	367.85	35.63	4.50×10^3	0.0	36.52	1.20
S11	367.85	35.73	1.00×10^6	0.0	37.09	1.70
S12	58.82	188.33	1.21×10^6	0.0	50.28	11.11
S13	18,636.00	15.00	1.01×10^5	0.0	0.00	0.00
S14	18,636.00	23.30	1.01×10^5	0.0	745.95	10.54
S15	410.75	323.36	3.65×10^6	1.0	1298.59	496.81
S20	367.85	360.50	1.03×10^6	1.0	1211.05	411.16
S21	410.75	538.00	4.00×10^6	1.0	1494.16	616.42
S22	15.98	423.23	1.72×10^6	1.0	54.54	20.02
S25	26.92	360.50	1.03×10^6	1.0	88.64	30.09
S33	309.62	35.63	4.50×10^3	0.93	774.70	54.07
S34	10.47	253.22	3.79×10^5	1.0	32.31	9.24
S35	23.88	209.93	2.41×10^5	1.0	71.73	18.82
S36	12.72	108.32	6.89×10^4	1.0	35.77	7.12
S37	11.16	60.47	3.45×10^4	1.0	30.40	5.03
S38	58.23	55.56	1.33×10^4	0.0	11.37	0.73
S39	367.85	124.86	1.00×10^6	0.0	195.94	30.41
S40	453.59	165.86	1.00×10^6	0.0	334.86	66.52
S41	453.59	169.28	1.62×10^7	0.0	347.05	77.57
S42	453.59	228.24	1.62×10^7	0.0	486.75	131.93

[a]The composition of all streams is 100% H_2O, except that, on a volume basis, the composition of S1 is 100% carbon, of S2 is 79% N_2 and 21% O_2, and of both S3 and S4 is 79% N_2, 6% O_2, and 15% CO_2.
[b]Vapor fraction is listed as 0.0 for liquids and 1.0 for superheated vapors.
[c]Stream S3 (not shown in Fig. 2) represents the hot product gases for adiabatic combustion.

section, in Table 2. Fig. 3a and b illustrate the net energy and exergy flows and exergy consumptions for the four main process sections.

Overall energy and exergy efficiencies are evaluated as

Energy efficiency

$$= (\text{Net energy output with electricity})/(\text{Energy input}) \quad (3)$$

and

Exergy efficiency

$$= (\text{Net exergy output with electricity})/(\text{Exergy input}) \quad (4)$$

Coal is the only input source of energy or exergy and the energy and exergy efficiencies are 37 and 36%,

Table 1b Data for principal thermal flows for a unit of the coal-fired electrical generating station

Stream	Energy flow rate (MW)	Exergy flow rate (MW)
Q5	5.34	0.00
Q6	5.29	0.00

Table 1c Data for principal electrical flows for a unit of the coal-fired electrical generating station

Stream	Energy (and exergy) flow rate (MW)
P1	0.57
P8	523.68
P15	12.19

Table 2 Breakdown of exergy consumption rates for a unit of the coal-fired electrical generating station

Section/device	Exergy consumption rate (MW)	
Steam generation		
Steam generator (including combustor)	659.0	
		659.0
Power production		
High-pressure turbine	26.4	
Intermediate-pressure turbine	22.3	
Low-pressure turbines	59.2	
Generator	5.3	
Transformer	5.3	
		118.5
Condensation		
Condenser	43.1	
		43.1
Preheat		
Low-pressure heat exchangers	10.7	
Deaerating heat exchanger	5.1	
High-pressure heat exchangers	6.4	
Hot well pumps	0.1	
Boiler feed pumps	1.1	
		23.4
Total		844.0

respectively. The small difference in the efficiencies is due to the fact that the specific chemical exergy of coal is slightly greater than its energy. Although the station energy and exergy efficiencies are similar, these efficiencies differ markedly for many station sections.

In the *Steam Generation* section, exergy consumptions are substantial, accounting for 659 MW (or 72%) of the 916 MW station exergy loss. Of this 659, 444 MW is consumed with combustion and 215 MW with heat transfer. The energy and exergy efficiencies for the *Steam Generation* section, considering the increase in energy or exergy of the water as the product, are 95 and 49%, respectively. The *Steam Generation* section thus appears significantly more efficient on an energy basis than on an exergy basis. Physically, this discrepancy implies that although 95% of the input energy is transferred to the preheated water, the energy is degraded as it is transferred. Exergy analysis highlights this degradation.

In the condensers, a large quantity of energy enters (775 MW for each unit), of which close to 100% is rejected, while a small quantity of exergy enters (54 MW for each unit), of which about 25% is rejected and 75% is internally consumed. Thus, energy analysis leads to the erroneous conclusion that almost all losses in electricity-generation potential for the station are associated with the heat rejected by the condensers, while exergy analysis demonstrates quantitatively and directly that the condensers are responsible for little of these losses (see Fig. 3b). This discrepancy arises because heat is rejected by the condensers at a temperature very near that of the environment.

In the *Power Production* and *Preheating* sections, energy losses are small (less than 10 MW) and exergy losses moderately small (118 MW in *Power Production* and 23 MW in *Preheating*). The exergy losses are almost completely associated with internal consumptions.

In assessing the thermodynamic characteristics of a coal-fired electrical generating station, several illuminating insights into performance are acquired:

- Although energy and exergy efficiencies are similar for the station, energy analysis does not identify the location and cause of process inefficiencies, while exergy analysis does. Energy losses are associated with emissions (mainly heat rejected by condensers) and exergy losses are primarily associated with consumptions (mainly in the combustors).
- Because devices with the largest thermodynamic losses have the largest margins for efficiency improvement, efforts to increase the efficiencies of coal-fired electrical generating stations should focus on the combustors. For instance, technologies capable of producing electricity without combustion (e.g., fuel cells) or utilizing heat at high temperatures could increase efficiencies significantly. This suggestion is, of course, overly simplistic, as such decisions must also account for other technical and economic factors.
- The use of heat rejected by condensers only increases the exergy efficiencies by a few percent.

APPLICATIONS BEYOND THERMODYNAMICS

Exergy concepts can be applied beyond thermodynamics in such fields as environmental impact assessment,[7,8,16] economics,[5,17,20,26] and policy.[27]

Exergy and Environment

Many suggest that the impact of energy utilization on the environment is best addressed by considering exergy. Although the exergy of an energy form or a substance is a measure of its usefulness, exergy is also a measure of its potential to cause change. The latter point suggests that exergy may be or may provide the basis for an effective

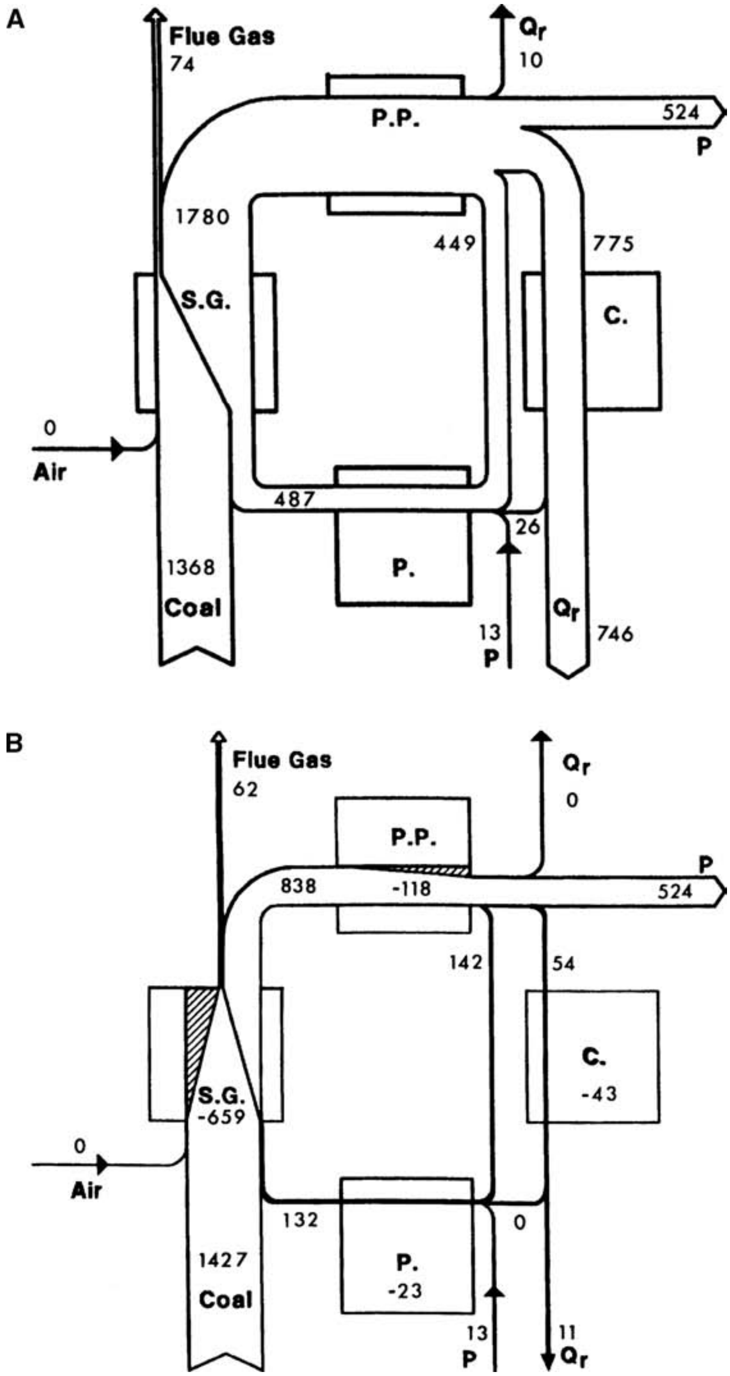


Fig. 3 (A) Diagram for a coal-fired electrical generating station unit indicating net energy flow rates (MW) for streams. Stream widths are proportional to energy flow rates. Station sections shown are *Steam Generation* (S.G.), *Power Production* (P.P.), *Condensation* (C.) and *Preheating* (P.). Streams shown are electrical power (P), heat input (Q), and heat rejected (Q_r). (B) Diagram for a coal-fired electrical generating station unit indicating net exergy flow rates for streams and consumption rates (negative values) for devices (in MW). Stream widths are proportional to exergy flow rates and shaded regions to exergy consumption rates. Other details are as in (A).

measure of the potential of a substance or energy form to impact the environment. The relation between exergy and the environment is discussed in this encyclopedia in an article entitled "Exergy: Environmental Impact Assessment Applications."

Exergy and Economics

Another area in which applications of exergy are increasing is that of economics. In the analysis and design of energy systems, techniques are often used that combine scientific disciplines like thermodynamics with economics to achieve optimum designs. For energy systems, costs are conventionally based on energy. Many researchers, however, have recommended that costs are better distributed among outputs based on exergy. Methods of performing exergy-based economic analyses have evolved (e.g., thermoeconomics, second-law costing, and exergoeconomics). These analysis techniques recognize that exergy, not energy, is the commodity of value in a system, and assign costs and prices to exergy-related variables. These techniques usually help in appropriately allocating economic resources so as to optimize the design and operation of a system and its economic feasibility and profitability (by obtaining actual costs of products and their appropriate prices).

CONCLUSION

Exergy analysis provides information that influences design, improvement, and application decisions and it is likely to be increasingly applied. Exergy also provides insights into the "best" directions for research, where "best" is loosely considered most promising for significant efficiency gains. There are two main reasons for this conclusion:

- Unlike energy losses, exergy losses represent true losses of the potential to generate the desired product from the given driving input. Focusing on exergy losses permits research to aim at reducing the losses that degrade efficiency.
- Unlike energy efficiencies, exergy efficiencies always provide a measure of how closely the operation of a system approaches the ideal. By focusing research on plant sections or processes with the lowest exergy efficiencies, effort is directed to those areas that inherently have the largest margins for efficiency improvement. By focusing on energy efficiencies, on the other hand, research can inadvertently be expended on areas for which little margins for improvement exist, even theoretically.

Exergy analysis results typically suggest that improvement efforts should concentrate more on internal rather than external exergy losses based on thermodynamic considerations, with a higher priority for the processes that have larger exergy losses. Of course, effort should still be devoted to processes having low exergy losses when cost-effective ways to increase efficiency can be identified.

Energy-related decisions should not be based exclusively on the results of energy and exergy analyses even though these results provide useful information to assist in such decision making. Other factors must also be considered, such as economics, environmental impact, safety, and social and political implications.

ACKNOWLEDGMENTS

The author is grateful for the support provided for this work by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Rezac, P.; Metghalchi, H. A brief note on the historical evolution and present state of exergy analysis. Int. J. Exergy **2004**, *1*, 426–437.
2. Kotas, T.J. *The Exergy Method of Thermal Plant Analysis*; Krieger: Malabar, FL, 1995; [Reprint Ed].
3. Moran, M.J. *Availability Analysis: A Guide to Efficient Energy Use*; American Society of Mechanical Engineers: New York, 1989; [Revised Ed].
4. Brodyanski, V.M.; Sorin, M.V.; Le Goff, P. *The Efficiency of Industrial Processes: Exergy Analysis and Optimization*; Elsevier: London, 1994.
5. Szargut, J.; Morris, D.R.; Steward, F.R. *Exergy Analysis of Thermal, Chemical and Metallurgical Processes*; Hemisphere: New York, 1988.
6. Yantovskii, E.I. *Energy and Exergy Currents (An Introduction to Exergonomics)*; Nova Science Publishers: New York, 1994.
7. Szargut, J. *Exergy Method: Technical and Ecological Applications*; WIT Press: Southampton, U.K., 2005.
8. Edgerton, R.H. *Available Energy and Environmental Economics*; D.C. Heath: Toronto, ON, 1982.
9. Moran, M.J.; Shapiro, H.N. *Fundamentals of Engineering Thermodynamics*, 5th Ed.; Wiley: New York, 2003.
10. Sato, N. *Chemical Energy and Exergy: An Introduction to Chemical Thermodynamics for Engineers*; Elsevier: Oxford, U.K., 2005.
11. Moran, M.J.; Sciubba, E. Exergy analysis: principles and practice. J. Eng. Gas Turbines Power **1994**, *116*, 285–290.
12. Rosen, M.A. Second law analysis: approaches and implications. Int. J. Energy Res. **1999**, *23*, 415–429.
13. Milia, D.; Sciubba, E. Exergy-based lumped simulation of complex systems: an interactive analysis tool. Energy **2006**, *31*, 100–111.
14. Frangopoulos, C.A.; von Spakovsky, M.R.; Sciubba, E. A brief review of methods for the design and synthesis optimization of energy systems. Int. J. Appl. Thermodyn. **2002**, *5* (4), 151–160.
15. Rosen, M.A.; Dincer, I. Effect of varying dead-state properties on energy and exergy analyses of thermal systems. Int. J. Therm. Sci. **2004**, *43* (2), 121–133.
16. Jorgensen, S.E.; Svirezhev, Y.M. *Towards a Thermodynamic Theory for Ecological Systems*; Elsevier: Oxford, U.K., 2004.
17. Rosen, M.A.; Dincer, I. Exergoeconomic analysis of power plants operating on various fuels. Appl. Therm. Eng. **2003**, *23* (6), 643–658.
18. Rosen, M.A.; Horazak, D.A. Energy and exergy analyses of PFBC power plants. In *Pressurized Fluidized Bed Combustion*; Alvarez Cuenca, M., Anthony, E.J., Eds., Chapman and Hall: London, 1995; 419–448.
19. Rosen, M.A.; Le, M.N.; Dincer, I. Efficiency analysis of a cogeneration and district energy system. Appl. Therm. Eng. **2005**, *25* (1), 147–159.
20. Rosen, M.A.; Dincer, I. Exergy-cost-energy-mass analysis of thermal systems and processes. Energy Convers. Manage. **2003**, *44* (10), 1633–1651.
21. Rosen, M.A.; Dincer, I. A study of industrial steam process heating through exergy analysis. Int. J. Energy Res. **2004**, *28* (10), 917–930.

22. Dincer, I.; Rosen, M.A. Energy and exergy analyses of thermal energy storage systems In *Thermal Energy Storage Systems and Applications*; Wiley: London, 2002; 411–510.
23. Rosen, M.A.; Tang, R.; Dincer, I. Effect of stratification on energy and exergy capacities in thermal storage systems. Int. J. Energy Res. **2004**, *28*, 177–193.
24. Rosen, M.A.; Dincer, I. Sectoral energy and exergy modeling Turkey. ASME J. Energy Resour. Technol. **1997**, *119*, 200–204.
25. Rosen, M.A. Evaluation of energy utilization efficiency in Canada using energy and exergy analyses. Energy Int. J. **1992**, *17*, 339–350.
26. Nuwayhid, R Y.; Jaber, M Y.; Rosen, M A.; Sassine, G P. On the thermodynamic treatment of diffusion-like economic commodity flows. Int. J. Exergy **2006**, *3* (1), 103–117.
27. Rosen, M.A. Energy-related education and awareness: the role and place for exergy in policy development. Int. J. Green Energy **2004**, *1* (4), 451–465.

Exergy: Environmental Impact Assessment Applications

Marc A. Rosen
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Exer–Fed

Abstract
Exergy can help to understand and mitigate environmental impact because exergy is a measure of the departure of a substance from equilibrium with a specified reference environment, which is often modeled as the actual environment. Furthermore, exergy is a measure of potential of a substance to cause change, so the exergy of an emission is a measure of the potential of the substance to change or impact the environment. Thus, exergy may be, or provide the basis for, an indicator of the potential of an emitted substance to impact the environment. This article describes exergy and its use via exergy analysis, the reference environment in exergy analysis, the relation between exergy and environmental impact, and how environmental impact assessment and reduction can be accomplished with exergy methods.

INTRODUCTION

Exergy analysis is a thermodynamic technique for assessing and improving systems and processes, which is similar but advantageous to energy analysis, in large part because it is based on the second law of thermodynamics. The exergy of an energy form or a substance is a measure of its usefulness.

Exergy can also be applied to understand and mitigate environmental impact because exergy is a measure of the departure of a substance from equilibrium with a specified reference environment, which is often modeled as the actual environment. Furthermore, exergy is a measure of potential of a substance to cause change. The exergy of an emission to the environment, therefore, is a type of measure of the potential of the substance to change or impact the environment. The greater the exergy of an emission, the greater is its departure from equilibrium with the environment, and the greater is its potential to change or impact the environment. The exergy of an emission is zero only when it is in equilibrium with the environment and thus benign. Consequently, exergy may be, or provide the basis for, an effective indicator of the potential of an emitted substance to impact the environment.

This article focuses on describing both the relations between exergy and environmental impact, and how environmental impact assessment can be accomplished with exergy methods. Such knowledge can permit the development of better understanding of, and more rational approaches to mitigating, the environmental impact associated with systems and processes. The scope of this article is as follows:

- Exergy and its use as an analysis technique via exergy analysis are briefly described, with particular attention on the role of the reference environment in exergy analysis.
- Relations between energy and the environment as well as exergy and the environment are discussed.
- Ways to reduce environmental impact with exergy are described.

Keywords: Exergy; Environmental impact; Emissions; Resource; Energy degradation.

EXERGY ANALYSIS

Exergy is defined as the maximum amount of work which can be produced by a flow of matter or energy as it comes to equilibrium with a reference environment. Exergy is a measure of the potential of a flow to cause change, as a consequence of being in disequilibrium with the reference environment.

The exergy Ex contained in a system may be written as

$$\text{Ex} = S(T - T_0) - V(p - p_0) + N_k(\mu_k - \mu_{k_0}) \qquad (1)$$

where the intensive properties are temperature, T, pressure, p, and chemical potential of substance k, μ_k; and the extensive properties are entropy, S, volume, V, and number of moles of substance k, N_k. The subscript "0" denotes conditions of the reference environment. It is evident from this expression that the exergy of a system is zero when it is in equilibrium with the reference environment (i.e., when $T = T_0$, $p = p_0$, and $\mu_k = \mu_{k_0}$ for all k). Clearly, evaluations of exergy require that the state of the reference environment, or the reference state, be specified, normally by its temperature, pressure, and chemical composition.

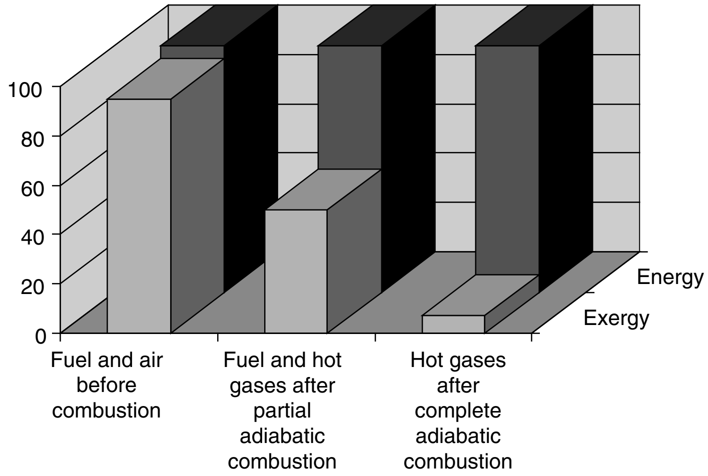


Fig. 1 Comparison of energy and exergy during combustion, for an input of 100 units of fuel energy.

Exergy analysis[1–7] uses the first and second laws of thermodynamics for the design and analysis of energy systems. Exergy is not subject to a conservation law, and is consumed due to the irreversibilities in any process. The exergy method is useful for attaining more efficient energy-resource use, for it enables the locations, types, and magnitudes of wastes and losses to be determined. Therefore, exergy analysis reveals by how much it is possible to design more efficient energy systems by reducing inefficiencies. Applications of exergy analysis occur in many areas and are increasing.

As a simple example, consider an adiabatic system containing fuel and air at ambient conditions. The fuel and air react to form a mixture of hot combustion gases. During the combustion process, the energy in the system remains fixed because it is adiabatic. But the exergy content declines as combustion proceeds due to the irreversibilities associated with the conversion of the high-quality energy of fuel to the lower quality energy of combustion gases. The different behaviors of energy and exergy during this process are illustrated qualitatively in Fig. 1.

More information on the fundamentals of exergy can be found in this encyclopedia in an article by the present author entitled "Exergy Analysis."

THE REFERENCE ENVIRONMENT IN EXERGY ANALYSIS

Exergy analysis requires a reference-environment model. Exergy is evaluated with respect to the reference-environment model, with the exergy of a flow and system being dependent on the intensive properties of the reference environment. The exergy of the reference environment is zero, and the exergy of a stream or system is zero when it is in equilibrium with the reference environment.

The reference environment has several theoretical characteristics. It is in stable equilibrium, and has all parts at rest relative to one another. In addition, the reference environment acts as an infinite system, is a sink and source for heat and materials, and experiences only internally reversible processes in which its intensive state remains unaltered (i.e., its temperature, pressure, and chemical potentials for each of the components present remain constant). No chemical reactions can occur between the reference-environment components.

The natural environment does not possess the theoretical characteristics of a reference environment, since it is not in equilibrium and its intensive properties vary spatially and temporally. Many chemical reactions in the natural environment are blocked because the transport mechanisms that are necessary to reach equilibrium are too slow at ambient conditions. Thus, the exergy of the natural environment is not zero, as work could be obtained if it were to come to equilibrium. In developing reference-environment models for exergy analysis, a compromise is made between the theoretical requirements of the reference environment and the behavior of the actual environment.

Several classes of reference-environment models are now described.

Natural-Environment-Subsystem Models

These models simulate subsystems of the natural environment. One such model consists of saturated moist air and liquid water in phase equilibrium.[8]

An extension of that model allows sulphur-containing materials to be analyzed.[9] The temperature and pressure of this reference environment (Table 1) are normally 25°C and 1 atm, respectively, and the chemical composition consists of air saturated with water vapor, and the following condensed phases at 25°C and 1 atm: water (H_2O), gypsum ($CaSO_4 \cdot 2H_2O$), and limestone ($CaCO_3$).

Table 1 A reference-environment model

Temperature, T_0	25°C
Pressure, P_0	1 atm
Composition	(i) Atmospheric air, saturated with H_2O at T_0 and P_0, having the following composition

Air constituents	Mole fraction
N_2	0.7567
O_2	0.2035
H_2O	0.0303
Ar	0.0091
CO_2	0.0003
H_2	0.0001

(ii) The following condensed phases at T_0 and P_0

Water (H_2O)

Limestone ($CaCO_3$)

Gypsum ($CaSO_4 \cdot 2H_2O$)

The stable configurations of C, O, and N, respectively, are taken to be those of CO_2, O_2, and N_2 as they exist in air saturated with liquid water at T_0 and P_0 (the temperature and pressure for the reference environment); of hydrogen is taken to be in the liquid phase of water saturated with air at T_0 and P_0; and of S and Ca, respectively, are taken to be those of $CaSO_4 \cdot 2H_2O$ and $CaCO_3$ at T_0 and P_0.

Reference-Substance Models

With this model, a "reference substance" is selected for every chemical element and assigned zero exergy. In one such model, the reference substances are selected as the most valueless substances found in abundance in the natural environment.[1,3] The criterion for selecting such reference substances is consistent with the notion of simulating the natural environment, but is primarily economic in nature, and is vague and arbitrary with respect to the selection of reference substances. Part of this environment is the composition of moist air, including N_2, O_2, CO_2, H_2O, and the noble gases; gypsum (for sulphur) and limestone (for calcium).

In a related model, reference substances are selected arbitrarily,[10] so the model does not resemble the natural environment. Absolute exergies evaluated with this model do not relate to the natural environment, and cannot be used rationally to evaluate efficiencies or environmental impact.

Equilibrium Models

In this model all the materials present in the atmosphere, oceans and a layer of the crust of the earth are pooled together and an equilibrium composition is calculated for a given temperature.[11] The selection of the thickness of crust considered is subjective and is intended to include all materials accessible to technical processes. Thicknesses varying from 1 to 1000 m, and a temperature of 25°C, have been considered. For all thicknesses, the model differs significantly from the natural environment. Exergy values obtained using these environments are significantly dependent on the thickness of crust considered, and represent the absolute maximum amount of work obtainable from a material. Since there is no technical process available which can obtain this work from materials, the equilibrium model does not give meaningful exergy values when applied to the analysis of real processes.

Constrained-Equilibrium Models

A modified version of the equilibrium environment model excludes in the equilibrium composition the possibility of the formation of nitric acid (HNO_3) and its compounds.[11] All chemical reactions in which these substances are formed are in constrained equilibrium, and all other reactions are in unconstrained equilibrium. When a thickness of crust of 1 m and temperature at 25°C are used, the model is similar to the natural environment.

Process-Dependent Models

This model contains only components that participate in the process being examined in a stable equilibrium composition at the temperature and pressure of the natural environment.[12] Being dependent on the process examined, the model is not general. Exergies evaluated for a specific process-dependent model are relevant only to the process, and cannot rationally be compared with the exergies evaluated for other process-dependent models or used in environmental assessments.

EXERGY, ENERGY AND THE ENVIRONMENT

Energy and the Environment

Energy production, transformation, transport, and end-use have significant impacts on the earth's environment. The present world distribution of per capita energy consumption suggests that a close correlation exists between a country's energy consumption and economic development. But, there does not appear to be a simple correlation between a country's energy consumption and the decay of its environment.

Exergy and the Environment

Some researchers suggest that exergy provides a less-subjective measure capable of providing greater insight into environmental impact.[13–19]

The exergy contents of waste emissions can be seen to be more meaningful than the corresponding energy contents as measures of potential for environmental impact. Material and energy flows only possess exergy when in disequilibrium with a reference environment. The exergy associated with waste emissions has the potential to cause environmental damage, particularly when released in an unconstrained manner into the environment. Many believe that by considering the exergy content of a waste emission, rational, and meaningful assessments can be made of the environmental impact potential of the emission.

Exergy losses, which consist of exergy destruction and waste exergy emissions, have a significant effect on environmental impact. Exergy destruction, in particular, is a significant criteria for assessing the depletion of natural resources. Exergy analysis can assist efforts to minimize the use of natural resources, as it indicates where the work potential of natural resources in relation to the surrounding environment is lost, i.e., where irreversibility, or exergy destruction, occurs.

A waste emission possesses exergy as a result of its being in a state of mechanical, thermal and/or chemical disequilibrium with the reference environment.[20,21] Exergy is often separated into two components: physical and chemical. From an environmental impact viewpoint, the exergy of an emission attributable to mechanical and thermal disequilibrium is not considered significant, as the potential impact on the environment of physical exergy is limited. That is, the pressure differences between an emission and the environment normally dissipate quickly and temperature differences are normally localized near the emission source (e.g., thermal pollution in the region of a lake near the cooling water discharge of a thermal power plant) and can be controlled. The exergy of an emission due to chemical disequilibrium (i.e., the chemical exergy), on the other hand, is often significant and not localized. The chemical exergy of emissions is normally a concern.

The exergy of a flow can only be entirely converted to products in a completely reversible process, since there are no exergy destructions within the system and no waste exergy emissions to the environment. In real-world processes, which are irreversible, both destruction of the exergy of resources and waste exergy emissions occur. Many efforts are being expended on reducing resource exergy destructions and eliminating waste exergy emissions—often by converting them into useful products. A reversible process represents a theoretical ideal which we can strive towards but never actually achieve.

REDUCING ENVIRONMENTAL IMPACT WITH EXERGY

Efforts to increase efficiency can reduce environmental impact by reducing exergy losses. Regional and national actions to improve exergy efficiency, in particular, can have major impacts on environmental protection. Work is being carried out on methods to reduce environmental impact with exergy methods. Some examples:

- Environomics is a methodology for analyzing and improving energy-related systems by simultaneously taking into account energy, exergy, economic, and environmental factors.[22] Exergy has also been integrated with life cycle analysis,[21,23] ecology and industrial ecology,[3,24–26] renewable concepts[27,28] and sustainability.[16,28–30]
- A recent study concluded that exergy analysis is a key tool for obtaining sustainable development. That study stated that, although energy can never be destroyed, exergy can be destroyed and this exergy destruction (irreversibility) must be appropriately minimized to obtain sustainable development. That study also showed that environmental effects associated with emissions and resource depletion can be expressed in terms of one indicator, which is based on physical principles.[17]
- The Consortium on Green Design and Manufacturing at the University of California-Berkeley is carrying out a project entitled "Exergy as an Environmental Indicator" to increase the practical application of exergy for rectifying problems associated with material and energy flows in industry.
- Several projects have recently been carried out at the Delft University of Technology and University of Twente to determine whether exergy analysis can be used in environmental policy development, especially for the comparison of alternative production chains. These studies are investigating if the linkages of exergy

with phenomena such as pollution and dispersion can be converted into a reliable tool on which policy decisions can be based, and are exploring how the environmental effects of processes can be linked to or expressed in terms of exergy changes.

Several researchers suggest the most appropriate way to link the second law and environmental impact is through exergy because it is a measure of the departure of the state of a system from that of the environment. The magnitude of the exergy of a system depends on the states of both the system and the environment. This departure is zero only when the system is in equilibrium with its environment. Tribus and McIrvine suggest that performing exergy analyses of the natural processes occurring on the earth could form a foundation for ecologically sound planning because it would indicate the disturbance caused by largescale changes.[18]

An understanding of the relations between exergy and the environment may reveal the fundamental patterns and forces affecting changes in the environment and help researchers deal better with environmental damage.

RELATIONS BETWEEN EXERGY AND ENVIRONMENTAL IMPACT

Three relations between exergy and environmental impact that help to explain patterns that underlie changes in the environment are now discussed.

Order Destruction and Chaos Creation

The destruction of order, or creation of chaos, is a form of environmental damage.

Fundamentally, entropy is a measure of chaos. A system of high entropy is more chaotic or disordered than one of low entropy. For example, a field with papers scattered about has a higher entropy than the field with the papers neatly piled. Conversely, exergy is a measure of order. Relative to the same environment, the exergy of an ordered system is greater than that of a chaotic one.

The difference between the exergy values of the two systems containing papers described above is a measure of the minimum work required to convert the chaotic system to the ordered one (i.e., in collecting the scattered papers). In reality, more than this minimum work, which only applies if a reversible clean-up process is employed, is required. The exergy destroyed when the wind scatters a stack of papers is a measure of the order destroyed during the process.

Resource Degradation

The degradation of resources found in nature is a form of environmental damage.

One definition of a resource is a material, found in nature or created artificially, in a state of disequilibrium with the environment. Resources have exergy as a consequence of this disequilibrium. For some resources (e.g., metal ores), it is their composition that is valued. For others (e.g., fuels), their reactivity is valued (i.e., their potential to cause change, or "drive" a task or process).

By preserving exergy through increased efficiency, i.e., degrading as little exergy as necessary for a process, environmental damage is reduced. Increased efficiency also has the effect of reducing exergy emissions.

The earth is an open system subject to a net influx of exergy from the sun. It is the exergy (or order states) delivered with solar radiation that is valued; all the energy received from the sun is ultimately radiated out to the universe. Environmental damage can be reduced by taking advantage of the openness of the earth and utilizing solar radiation, instead of degrading resources found in nature to supply exergy demands. This would not be possible if the earth was a closed system, for it would eventually become increasingly degraded, or "entropic."

Waste Exergy Emissions

The exergy associated with process wastes emitted to the environment can be viewed as a potential for environmental damage. (This should not be confused with exergy in the environment that is constrained and thus a resource, as shown in Fig. 2).

Typical process wastes have exergy, a potential to cause change, as a consequence of not being in stable equilibrium with the environment. When emitted to the environment, this exergy represents a potential to change the environment. In some cases, this exergy may cause a change perceived to be beneficial (e.g., the increased rate of growth of fish and plants near the cooling-water outlets from thermal power plants). More often, however, emitted exergy causes a change which is damaging to the

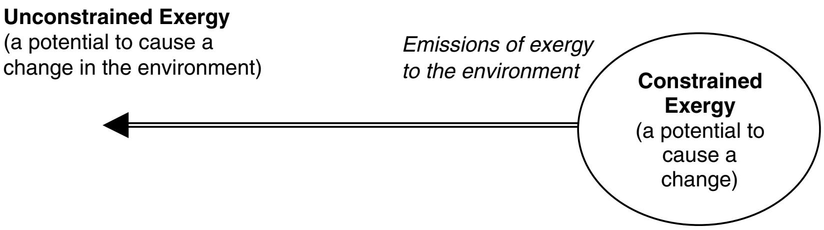


Fig. 2 Comparison of constrained and unconstrained exergy. Exergy constrained in a system represents a resource, while exergy emitted to the environment becomes unconstrained and represents a potential for environmental damage.

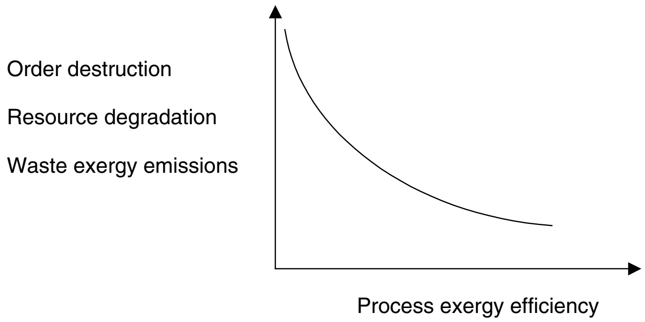


Fig. 3 Qualitative illustration of the relations between the exergy efficiency of a process and the associated environmental impact in terms of order destruction, resource degradation and waste exergy emissions.

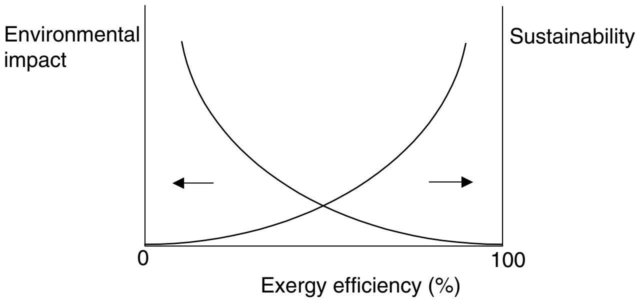


Fig. 4 Qualitative illustration of the relations between the environmental impact and sustainability of a process, and its exergy efficiency.

environment (e.g., the deaths of fish and plants in some lakes due to the release of specific substances in stack gases as they react and come to equilibrium with the environment).

Emissions of exergy to the environment can also interfere with the net input of exergy via solar radiation to the earth. Carbon dioxide emissions appear to be changing the atmospheric CO_2 concentration, affecting the receiving and re-radiating of solar radiation by the earth.

The relation between waste exergy emissions and environmental damage has been recognized by several researchers. By considering the economic value of exergy in fuels, one air-pollution rating estimates the air-pollution cost for a fuel as either the cost to remove the pollutant or the cost to society of the pollution (i.e., the tax which should be levied if pollutants are not removed from effluent streams).[2] Reistad claims the rating is preferable to mainly empirical ratings commonly used.

Discussion

The decrease in the environmental impact of a process, in terms of the three measures discussed in this section, as exergy efficiency increases for a process is illustrated in Fig. 3. Fig. 3 can be expanded as shown in Fig. 4 to illustrate the tie to sustainability. There sustainability is seen to increase and environmental impact to decrease as process exergy efficiency increases. Two limiting cases in Fig. 4 are significant:

- As exergy efficiency approaches 100%, environmental impact approaches zero, since exergy is converted from one form to another without loss (internal consumptions or waste emissions), and sustainability approaches infinity because the process approaches reversibility.
- As exergy efficiency approaches 0%, sustainability approaches zero because exergy-containing resources are used but nothing is accomplished, and environmental impact approaches infinity because to provide a fixed service, an ever-increasing quantity of resources must be used and a correspondingly increasing amount of exergy-containing wastes are emitted.

CONCLUSION

The relations between environmental impact and exergy could lead to simple, rational, and objective procedures for assessing the harmful effects on the environment and predicting the potential for environmental impact for a substance. This view is primarily based on the general premise that a substance has exergy due to its disequilibrium with respect to a reference environment, and therefore the exergy of unconstrained waste emissions has the potential to impact on the environment. The potential usefulness of exergy analysis in addressing and solving environmental problems is substantial, although further work is needed if utilizable exergy-based environmental-impact measures and methods are to be developed.

ACKNOWLEDGMENTS

The author is grateful for the support provided for this work by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Szargut, J.; Morris, D.R.; Steward, F.R. *Exergy Analysis of Thermal, Chemical and Metallurgical Processes*; Hemisphere: New York, U.S., 1988.
2. Reistad, G.M. *Availability: Concepts and Applications*, Ph.D. dissertation, University of Wisconsin: Madison, 1970.
3. Szargut, J. *Exergy Method: Technical and Ecological Applications*; WIT Press: Southampton, U.K., 2005.

4. Sato, N. *Chemical Energy and Exergy: An Introduction to Chemical Thermodynamics for Engineers*; Elsevier: Oxford, U.K., 2005.
5. Kotas, T.J. *The Exergy Method of Thermal Plant Analysis*, Reprint Ed.; Krieger: Malabar, FL, 1995.
6. Moran, M.J. *Availability Analysis: A Guide to Efficient Energy Use*, Revised Ed.; American Society of Mechanical Engineers: New York, 1989.
7. Edgerton, R.H. *Available Energy and Environmental Economics*; D.C. Heath: Toronto, 1982.
8. Baehr, H.D.; Schmidt, E.F. Definition und berechnung von brennstoffexergien (Definition and calculation of fuel exergy). Brennst-Waerme-Kraft **1963**, *15*, 375–381.
9. Gaggioli, R.A.; Petit, P.J. Use the second law first. Chem. tech. **1977**, *7*, 496–506.
10. Sussman, M.V. Steady-flow availability and the standard chemical availability. Energy Int. J. **1980**, *5*, 793–804.
11. Ahrendts, J. Reference states. Energy Int. J. **1980**, *5*, 667–678.
12. Bosnjakovic, F. Bezugszustand der exergie eines reagiernden systems (Reference states of the exergy in a reacting system). Forsch. Ingenieurw. **1963**, *20*, 151–152.
13. Creyts J.C.; Carey, V.P. Use of extended exergy analysis as a tool for assessment of the environmental impact of industrial processes. In Proceedings of the ASME Advanced Energy Systems Division, Ramalingam, M.L., Lage, J.L., Mei, V.C., Chapman, J.N., Eds.; American Society of Mechanical Engineers: New York, 1997; AES-Vol. 37, 129–137.
14. Gunnewiek, L.H.; Rosen, M.A. Relation between the exergy of waste emissions and measures of environmental impact. Int. J. Environ. Pollut. **1998**, *10* (2), 261–272.
15. Rosen, M.A.; Dincer, I. On exergy and environmental impact. Int. J. Energy Res. **1997**, *21*, 643–654.
16. Rosen, M.A.; Dincer, I. Exergy as the confluence of energy, environment and sustainable development. Int. J. Exergy **2001**, *1*, 3–13.
17. Sciubba, E. Exergy as a direct measure of environmental impact. In Proceedings of the ASME Advanced Energy Systems Division, Aceves, S.M., Garimella, S., Peterson, R., Eds.; American Society of Mechanical Engineers, New York, 1999; AES-Vol. 39, 573 581.
18. Tribus, M.; McIrvine, E.C. Energy and information. Sci. Am. **1971**, *225* (3), 179–188.
19. Rosen, M.A. Can exergy help us understand and address environmental concerns? Int. J. Exergy **2002**, *2* (4), 214–217.
20. Rosen, M.A.; Dincer, I. Exergy analysis of waste emissions. Int. J. Energy Res. **1999**, *23*, 1153–1163.
21. Daniel, J.J.; Rosen, M.A. Exergetic environmental assessment of life cycle emissions for various automobiles and fuels. Int. J. Exergy **2002**, *2* (4), 283–294.
22. Pelster, S.; Favrat, D.; von Spakovsky, M.R. The thermoeconomic and environomic modeling and optimization of the synthesis, design and operation of combined cycles with advanced options. Trans. ASME J. Eng. Gas Turbines Power **2001**, *123* (4), 717–726.
23. Ayres, R.U.; Ayres, L.W.; Martinas, K. Exergy, waste accounting, and life-cycle analysis. Int. J. Energy **1998**, *23*, 355–363.
24. Connelly, L.; Koshland, C.P. Exergy and industrial ecology. Int. J. Exergy **2001**, *1*, 146–165 and 234–255.
25. Jorgensen, S.E.; Svirezhev, Y.M. *Towards a Thermodynamic Theory for Ecological Systems*; Elsevier: Oxford, U.K., 2004.
26. Salas, F.; Marcos, C.; Pérez-Ruzafa, A.; Marque, J.C. Application of the exergy index as ecological indicator of organically enrichment areas in the Mar Menor lagoon (South-Eastern Spain). Energy **2005**, *30*, 2505–2522.
27. Berthiaume, R.; Bouchard, C.; Rosen, M.A. Exergetic evaluation of the renewability of a biofuel. Int. J. Exergy **2001**, *1*, 256–268.
28. Dincer, I.; Rosen, M.A. Thermodynamic aspects of renewables and sustainable development. Renew. Sust. Energy Rev. **2005**, *9* (2), 169–189.
29. Wall, G.; Gong, M. On exergy and sustainable development. Int. J. Exergy **2001**, *1*, 128–145 and 217–233.
30. Dincer, I.; Rosen, M.A. Exergy as a driver for achieving sustainability. Int. J. Green Energy **2004**, *1* (1), 1–19.

Facility Air Leakage

Wendell A. Porter
University of Florida, Gainesville, Florida, U.S.A.

Abstract
Building air leakage issues are described in detail in this entry. The advantages and disadvantages of the most popular techniques and materials used to reduce air leakage are discussed. Information concerning the most often missed avenues for air leakage through penetrations and air flow bypasses is also provided.

Each major building component that contributes to air leakage, such as floors, ceilings and walls, is described in detail. The proper application method for each technique used to reduce air leakage is explained in both the text and illustrations. Measurement techniques utilizing blower door technology are also introduced. These techniques have been used to measure the success of the goal of reducing unintended building air leakage.

INTRODUCTION

Air leakage is a major problem for both new and existing buildings, and it can:

- Contribute to over 30% of cooling and heating costs
- Create comfort and moisture problems
- Pull pollutants, such as radon and mold, into buildings
- Serve as easy access for insects and rodents

To effectively reduce air leakage requires a continuous air barrier system—a combination of materials linked together to create a tight building envelope (Fig. 1). An effective building envelope should form both a continuous air barrier and an insulation barrier. An air barrier minimizes air currents inside the cavities of the building envelope, which helps maintain insulation *R*-values.

The air barrier should seal all leaks through the building envelope—the boundary between the conditioned portion of the building and the unconditioned area. Most standard insulation products are not effective at sealing air leakage. The *R*-value for these products may drop if air leaks through the material.

The builder should work with his or her own crew and subcontractors to seal all penetrations through the envelope. Then, continuous material should be installed around the envelope. In the air sealing process, it is critical to use durable materials and install them properly.

INFILTRATION CONTROL

Vapor Retarders

An infiltration barrier shall provide a continuous air barrier from the foundation to the top plate of the ceiling and shall be sealed at the foundation, the top plate, at openings in the wall plane (windows, doors, etc.), and at the seams between sections of infiltration barrier material. When installed on the interior side of the walls, such as with insulated face panels with an infiltration barrier, the infiltration barrier should be sealed at the foundation or subfloor. This prevents wind from circulating air within the insulation. If properly sealed at the seams and ends, plywood and builder's felt will serve as an effective infiltration barrier, but not as a moisture retardant. A vapor retarder will essentially stop moisture transmission or diffusion. Common vapor retarders include 6-mil polyethylene sheet and aluminum foil-backed paper or boards. Contrary to northern construction practices, a vapor retarder, including vinyl wall coverings, installed next to the conditioned space is not recommended; during the summer, water may condense on the vapor retarder surface within the wall cavity when the inside temperature is below the outside dew point. This could wet and degrade insulation, deteriorate wall components, and contribute to mold and mildew. Vapor retarders are not recommended on the conditioned side of walls in Florida buildings. In fact, the ASHRAE 2001 Fundamentals Handbook does not recommend vapor retarders at all in hot, humid climates (on walls or ceilings). It should be noted that vinyl wallpaper acts like a vapor barrier. Placement on the conditioned side of the wall may cause moisture problems.

Air Barriers

Housewraps serve as exterior air barriers and help reduce air leakage through outside walls. Most products block only air leakage, not vapor diffusion, so they are not vapor retarders.

Typical products are rolled sheet materials that can be affixed and sealed to the wall between the sheathing and exterior finish material (Fig. 2). For best performance, a housewrap must be sealed with caulk or tape at the top and bottom of the wall and around any openings, such as

Keywords: Air leakage; Vapor barrier; Continuous air barrier; Insulation barrier; Infiltration.

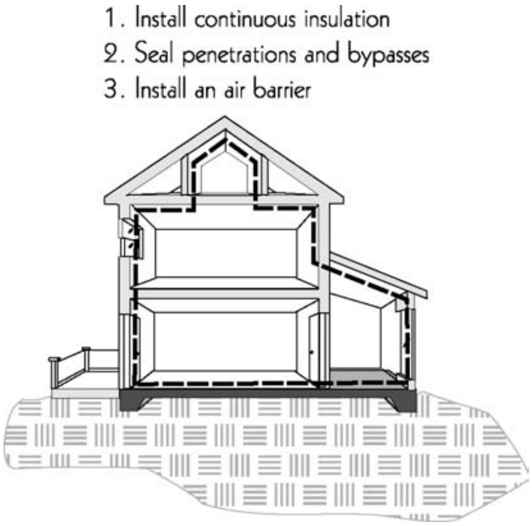


Fig. 1 Tight building envelope.

windows, doors, and utility penetrations, and be installed per manufacturer's specifications.

A housewrap can help reduce air leakage through exterior walls, but, by itself, is not a continuous air barrier for the entire envelope and, hence, is not a substitute for the airtight drywall approach. Housewraps are primarily recommended as further insurance against air leakage and, because they can block liquid water (bulk water) penetration, can help protect a building from moisture damage. In some instances, the exterior sheathing may be used as an outside air barrier (Fig. 3). Careful sealing of all seams and penetrations, including windows, is required (Fig. 4).

Materials

Most air barrier systems rely on a variety of caulks, gaskets, weatherstripping, and sheet materials such as plywood, drywall, and housewraps.

Use a combination of these different air sealing materials:

- Caulk—Use to seal gaps less than 1/4 in. Select the grade (interior, exterior, high temperature) based on application.
- Spray foam—Expands to fill large cracks and small holes. It can be messy; consider new, water-based foams. Not recommended near flammable applications (flue vents, etc.) and not permitted around PVC pipe. May be prohibited around windows by window manufacturers.
- Gaskets—Apply under the bottom plate before an exterior wall is raised or used to seal drywall to framing.
- Housewrap—Installed between the sheathing and exterior finish material. Must be sealed with tape or caulk to form an airtight seal. Resists liquid water and is not a vapor retarder.
- Sheet goods (plywood, drywall, rigid foam insulation)—These are the solid materials that form the building envelope. Air will only leak at the seams or through unsealed penetrations.
- Sheet metal—Used with high temperature caulk for sealing high temperature components, such as flues, to framing.
- Polyethylene plastic—Inexpensive material for air-sealing that also stops vapor diffusion. Must have all

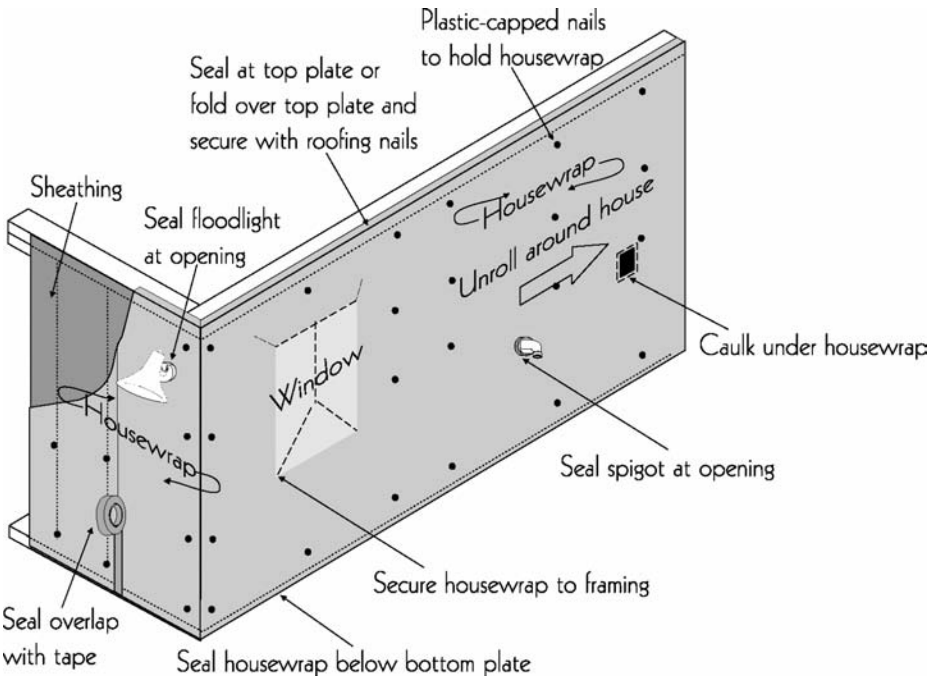


Fig. 2 Housewrap air barrier.

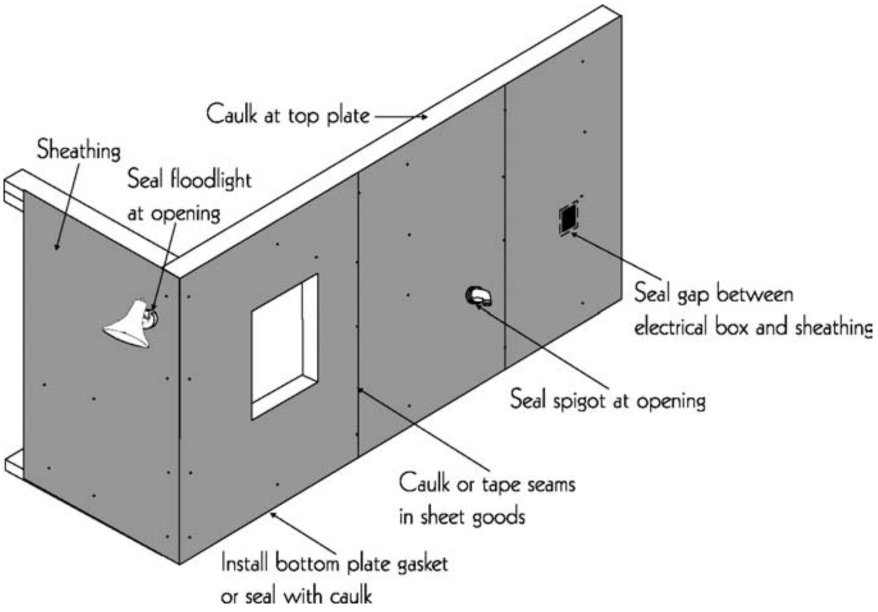


Fig. 3 Exterior sheathing air barrier.

edges and penetrations sealed to be effective air barriers. This material is recommended for use under slabs. Incorrect application in wall cavities can lead to mold and mildew problems.

- Weatherstripping—Used to seal moveable components, such as doors and windows.

Seal Penetrations and Bypasses

The first step in successfully creating an air barrier system is to seal all of the holes in the building envelope. Too often, builders concentrate on air leakage through windows, doors, and walls, and ignore areas of much greater importance. Many of the key sources of leakage—called bypasses (Fig. 5)—are hidden from view behind soffits for cabinets, bath fixtures, dropped ceilings, chases for flues and ductwork, recessed lighting fixtures, or insulation. Attic access openings and whole house fans are also common bypasses. Sealing these bypasses is crucial in reducing a building's air leakage and maintaining the performance of insulation materials. Table 1 provides examples of commonly used sealants for various types of leaks.

Air Leakage Details

The guidelines that follow in Fig. 6 show important areas that should be sealed to create an effective air barrier. The builder must clearly inform subcontractors and workers of these details to ensure that the task is accomplished successfully.

- Slab Floors—If a house is to be constructed on a concrete slab, a vapor retarder of plastic sheeting should be placed under the slab. Without a vapor retarder, moisture will migrate from the ground through the porous slab and into the house. If a house is to be built off-grade, a sheet of 6-mil polyethylene plastic should always be placed directly on the ground

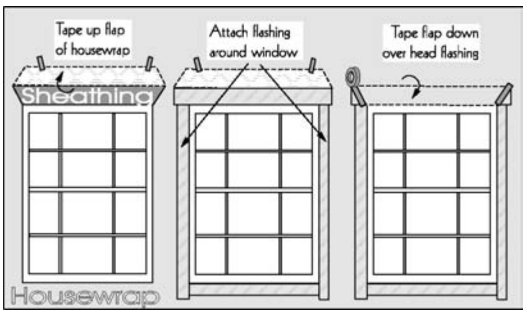


Fig. 4 Air barrier sealing around windows.

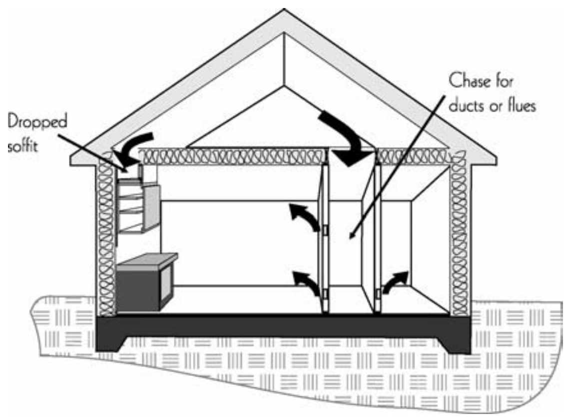


Fig. 5 Air leakage bypasses.

Table 1 Leaks and sealants

Type of leak	Commonly used sealants
Thin gaps between framing and wiring, pipes, or ducts through floors or walls	40-Year caulking; one-part polyurethane is recommended
Leaks into attics, cathedral ceilings, and wall cavities above first floor	Firestop caulking, foam sealant (latex or polyurethane)
Gaps, or cracks or holes over 1/8 in. in width that do not require firestop sealant	Gasket, foam sealant, or stuff with fiber glass or backer rod, and caulk on top
Open areas around flues, chases, plenums, plumbing traps, etc.	Attach and caulk a piece of plywood or foam sheathing material that covers the entire opening
	Seal penetrations
	If a flue requires a non-combustible clearance, use a noncombustible metal collar, sealed in place, to span the gap
Final air barrier material system	Install airtight drywall approach or other air barrier

under the house to prevent moisture from moving upward from the soil.

- Floor Joist—Seal sill plates in basements and unvented crawl spaces. Caulk or gasket rim or band joists between floors in multi-story construction.
- Bottom Plate—Use either caulk or gasket between the plate and subflooring.
- Electrical Wiring—Use wire-compatible caulk or spray foam to seal penetrations.
- Electrical Boxes—Use approved caulk to seal wiring on the outside of electrical boxes. Seal between the interior finish material and boxes.

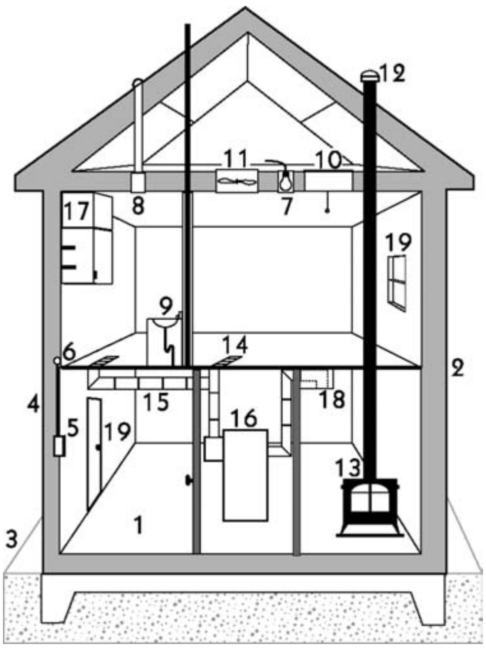


Fig. 6 Air leakage details.

- Electrical Box Gaskets—Caulk foam gaskets to all electrical boxes in exterior and interior walls before installing coverplates.
- Recessed Light Fixtures—Consider using surface-mounted light fixtures rather than recessed lights. When used, specify airtight models rated for insulation contact. Ensure fixtures meet appropriate fire codes.
- Exhaust Fans—Seal between the fan housing and the interior finish material. Choose products with tight-fitting backdraft dampers.
- Plumbing—Locate plumbing in interior walls, and minimize penetrations. Seal all penetrations with sealant or caulk.
- Attic Access in Conditioned Spaces—Weatherstrip attic access openings. For pull-down stairs, use latches to hold the door panel tightly against the weatherstripping. Cover the attic access opening with an insulated box.
- Whole House Fan—Use a panel made of rigid insulation or plastic to seal the interior louvers.
- Flue Stacks—Install a code-approved flue collar and seal with fire-rated caulk (Fig. 7).
- Combustion Appliances—Closely follow all codes for firestopping measures, which reduce air leakage as well as increase the safety of the appliance. Make certain all combustion appliances, such as stoves, inserts, and fireplaces, have an outside source of combustion air and tight-fitting dampers or doors.
- Return and Supply Registers—Seal all boots connected to registers or grilles to the interior finish material.
- Ductwork—Seal all joints in supply and return duct systems with mastic. Mechanically attach duct systems to prevent dislocation and massive leakage.
- Air Handling Unit (for heating and cooling system)—Seal all cracks and unnecessary openings with mastic. Seal service panels with UL 181 listed and labeled tape.

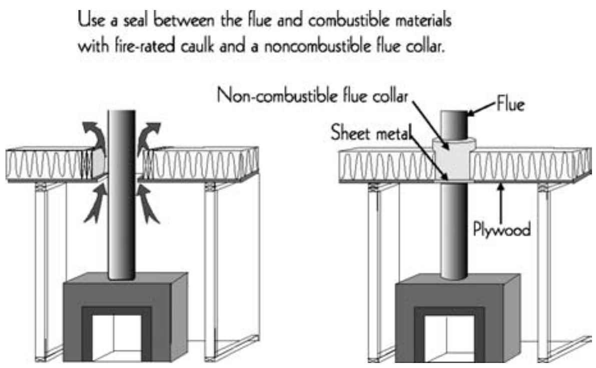


Fig. 7 Flue stacks air leakage details.

- Dropped Ceiling Soffit—Use sheet material and sealant to stop air leakage from attic into the soffit or wall framing, then insulate (see Figs. 4–8 for more details).
- Chases (for ductwork, flues, etc.)—Prevent air leakage through these bypasses with sheet materials and sealants (see Figs. 7 and 9 for more details).
- Windows and Doors—Must meet allowable air infiltration rates found in state building codes.

Air Leakage Driving Forces

Requirements for air leakage to occur:

- Holes—The larger the hole, the greater the air leakage. Large holes have higher priority for air sealing efforts.
- Driving Force—A pressure difference that forces air to flow through a hole. Holes that experience stronger and more continuous driving forces have higher priority.

The common driving forces are:

- Wind—Caused by weather conditions.
- Mechanical Blower—Induced pressure imbalances caused by the operation of fans and blowers.
- Stack Effect—Upward air pressure due to the buoyancy of air.

Wind is usually considered the primary driving force for air leakage. When the wind blows against a building, it creates a high pressure zone on the windward areas. Outdoor air from the windward side infiltrates into the building, while air exits on the leeward side. Wind acts to create areas of differential pressure, which cause both infiltration and exfiltration. The degree to which wind contributes to air leakage depends on its velocity and duration. In the following Figs. 10–12, a "Pascal" is the metric system unit of measure for stress or force per unit area.

Poorly designed and installed forced-air systems can create strong pressure imbalances inside buildings, which can triple air leakage whenever the heating and cooling system operates. In addition, unsealed ductwork located in attics and crawl spaces can draw pollutants and excess moisture into the building. Correcting duct leakage problems is critical when constructing an energy efficient building.

The temperature difference between inside and outside causes warm air inside the building to rise while cooler air falls, creating a driving force known as the stack effect. The stack effect is weak, but always present. Most buildings have large holes into the attic and crawl space. Because the stack effect is so prevalent and the holes through which it drives air are often so large, it is usually a major contributor to air leakage, moisture, and air quality problems.

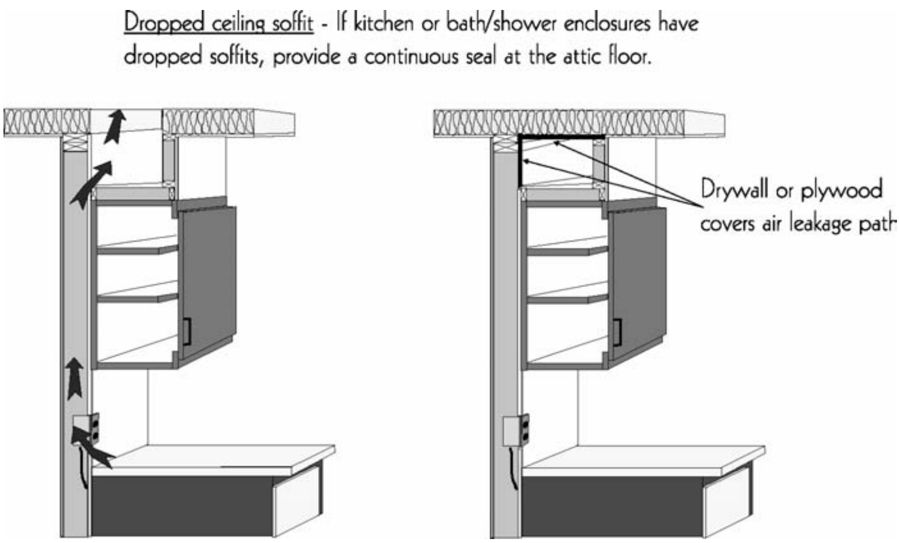


Fig. 8 Dropped ceiling soffit air leakage details.

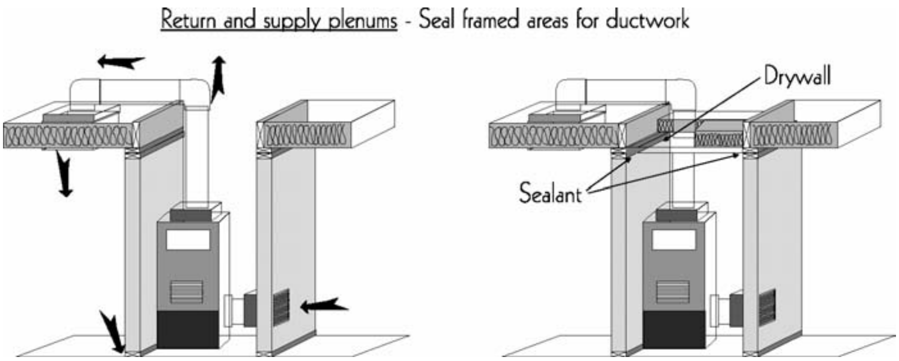


Fig. 9 Flue stacks air leakage details.

Airtight Drywall Approach

The airtight drywall approach is an air sealing system that connects the interior finish of drywall and other building materials together to form a continuous barrier (Fig. 13). The airtight drywall approach has been used on hundreds of houses and has proven to be an effective technique to reduce air leakage, as well as keep moisture, dust, and insects from entering the building.

In a typical drywall installation, most of the seams are sealed by tape and joint compound. However, air can leak in or out of the building in the following locations:

- Between the edges of the drywall and the top and bottom plates of exterior walls.
- From inside the attic down between the framing and drywall of partition walls.
- Between the window and door frames and drywall.
- Through openings in the drywall for utilities and other services.

The airtight drywall approach uses either caulk or gaskets to seal these areas and make the drywall a continuous air barrier system.

Advantages

- Effective—The airtight drywall approach has proven to be a reliable air barrier.
- Simple—Does not require specialized subcontractors or unusual construction techniques. If gasket materials are not available locally, they can be shipped easily.
- Does Not Cover Framing—The use of the airtight drywall approach does not prevent drywall from being glued to the framing.
- Scheduling—Gaskets can be installed anytime between when the house is "dried-in" and when the drywall is attached to framing.
- Adaptable—Builders can adapt airtight drywall approach principles to suit any design and varying construction schedules.
- Cost—Materials and labor for standard designs should only cost a few hundred dollars.

Disadvantages

- New—Although the airtight drywall approach is a proven technique, many building professionals and code officials are not familiar with its use.
- Requires Thought—While the airtight drywall approach is simple, new construction techniques require careful planning to ensure that the air barrier remains continuous. However, the airtight drywall approach is often the most error-free and reliable air barrier for unique designs.
- Requires Care—Gaskets and caulking can be damaged or removed by subcontractors when installing the drywall or utilities.

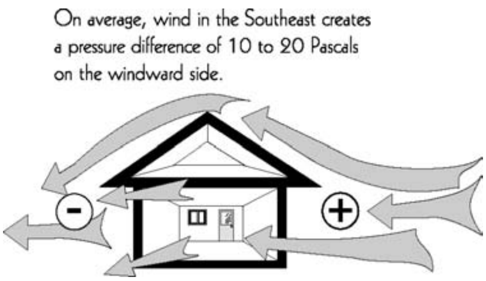


Fig. 10 Wind effect on houses.

Fig. 11 HVAC equipment effect on building pressure differences.

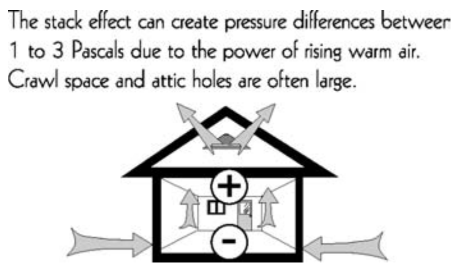


Fig. 12 Stack effect on building pressure differences.

Installation Techniques

Slab Floors

- Seal expansion joints and penetrations with a concrete sealant, such as one-part urethane caulk.

Exterior Framed Walls

- Seal between the bottom plate and subflooring with caulk or gaskets.
- Install gaskets or caulk along the face of the bottom plate so that when drywall is installed, it compresses the sealant to form an airtight seal against the framing. Some builders also caulk the drywall to the top plate to reduce leakage into the wall.
- Use drywall joint compound or caulk to seal the gap between drywall and electrical boxes. Install foam gaskets behind coverplates and caulk holes in boxes.
- Seal penetrations through the top and bottom plates for plumbing, wiring, and ducts. The code requires firestopping for leaks through top plates.

Partition Walls

- Seal the drywall to the top plate of partition walls with unconditioned space above.

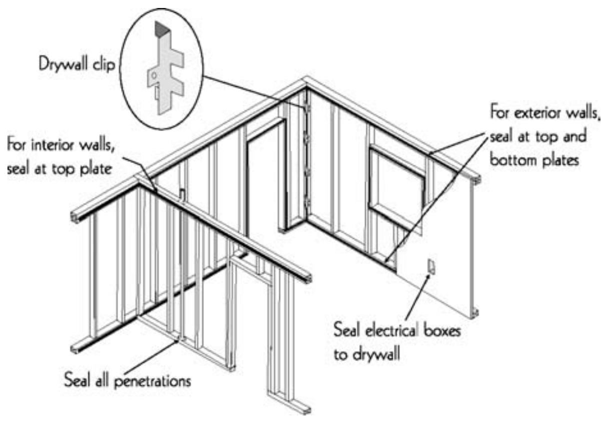


Fig. 13 Airtight drywall approach.

- Install gaskets or caulk on the face of the first stud in the partition wall. Sealant should extend from the bottom to the top of the stud to keep air in the outside wall from leaking inside.
- Seal the ductwork where it projects through partition walls.
- Seal penetrations through the top and bottom plates for plumbing, wiring, and ducts. Fire code requirements dictate how this should be accomplished.

Windows and Doors

- Seal drywall edges to either framing or jambs for windows and doors.
- Fill rough opening with spray foam sealant or suitable substitute.
- Caulk window and door trim to drywall with clear or paintable sealant.

Ceiling

- Follow standard finishing techniques to seal the junction between the ceiling and walls.
- When installing ceiling drywall, do not damage gaskets, especially in tight areas such as closets and hallways.
- Seal all penetrations in the ceiling for wiring, plumbing, ducts, attic access openings, and whole house fans.
- Seal all openings for chases and dropped soffits above kitchen cabinets and shower/tub enclosures.
- Avoid recessed lights; where used, install airtight, IC-rated fixtures and caulk between fixtures and drywall.

Wood Framed Floors

- Seal the rim joist to minimize air currents around floor insulation. Also, seal rim joists for multi-story construction (Fig. 14).
- For unvented crawl spaces or basements, seal beneath the sill plate.
- Seal the seams between pieces of subflooring with quality adhesive.

MEASURING AIRTIGHTNESS WITH A BLOWER DOOR

While there are many well-known sources of air leakage, virtually all buildings have unexpected air leakage sites called bypasses. These areas can be difficult to find and correct without the use of a blower door. This diagnostic equipment consists of a temporary door covering, which is installed in an outside doorway, and a fan that pressurizes

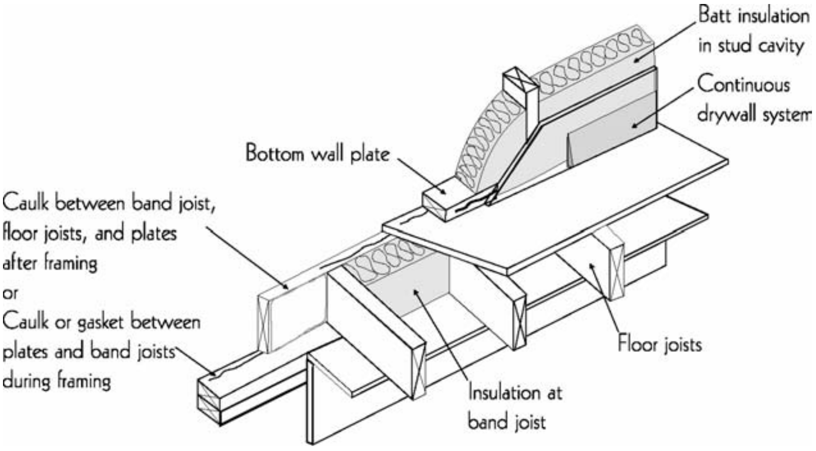


Fig. 14 Between-floor air barrier.

(forces air into) or depressurizes (forces air out of) the building. When the fan operates, it is easy to feel air leaking through cracks in the building envelope. Most blower doors have gauges which can measure the relative leakiness of a building (Fig. 15).

One measure of a building's leakage rate is air changes per hour (ACH), which estimates how many times in one hour the entire volume of air inside the building leaks to the outside. For example, a home that has 2000 square feet of living area and 8-foot ceilings has a volume of 16,000 cubic feet. If the blower door measures leakage of 80,000 cubic feet per hour, the home has an infiltration rate of 5 ACH. The leakier the house, the higher the number of air changes per hour, the higher the heating and cooling costs, and the greater the potential for moisture, comfort, and health problems (Table 2).

To determine the number of air changes per hour, many experts use the blower door to create a negative pressure of 50 Pascals. A Pascal is a small unit of pressure about equal to the pressure that a pat of butter exerts on a piece of toast—about 0.004 in. water gauge. Fifty Pascals is approximately equivalent to a 20 mile-per-hour wind blowing against all surfaces of the building. Energy efficient builders should strive for less than five air changes per hour at 50 Pascals pressure (ACH50). Given ACH50, a natural infiltration rate (resulting from wind and temperature effects) can be estimated. In hot, humid climates, ACH50 can be divided by 40 to yield an expected natural infiltration rate. For example, if ACH50 = 10, then the estimate for natural infiltration for hot, humid climate homes would be 10/40 = 0.25 ach. This means that under normal wind and temperature conditions, we would expect about 25% of the house air to be replaced with outdoor air each hour. Note that this is only an estimate of long-term average infiltration. Actual infiltration will vary considerably based on changes in wind, temperature, and time of day.

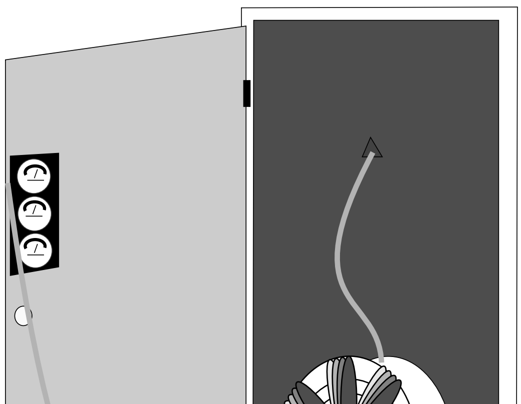


Fig. 15 Blower door test equipment.

CONCLUSION

Building air leakage or infiltration can be a significant contributor to increased energy costs, comfort issues, and unacceptable levels of indoor pollution. However, with the incorporation of proper materials and techniques,

Table 2 Typical infiltration rates for homes (in air changes per hour at 50 Pascals—ACH50)

New home with special airtight construction and a controlled ventilation system	1.5–2.5
Energy efficient home with continuous air barrier system	4.0–6.0
Standard new home	7.0–15.0
Standard existing home	10.0–25.0
Older, leaky home	20.0–50.0

problems associated with air leakage can be reduced. The reduction of a building's total air leakage requires an overall strategy coupled with an appreciation of the techniques and materials required. Proper implementation of these techniques can result in significant reduction in a building's energy consumption coupled with a positive economic picture.

Much attention has been given to the proper use of air infiltration barriers or housewraps and their success at reducing unwanted infiltration. More attention should be directed to building penetrations and airflow bypasses. Construction techniques need to be modified to reduce unintended air infiltration that occurs as a consequence of current construction practices.

The growing popularity of blower door technology has allowed the industry to track progress toward reducing building air leakage. Progress has been significant, even to the point of reducing infiltration below the amount needed for fresh air requirements. Reductions to this level allow mechanical systems designers to implement active ventilation schemes that both enhance healthy conditions and reduce energy consumption.

ACKNOWLEDGMENT

Many thanks to Kyle Allen, a mechanical engineering student at the University of Florida, because without his persistent efforts the publication of this article would not have been possible.

BIBLIOGRAPHY

1. ASHRAE Handbook—Fundamentals. American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. (ASHRAE), 1791 Tullie Circle, NE, Atlanta, GA 30329-2305. May be ordered online at: http://www.ashrae.org.
2. Cummings, J.B.; Tooley, J.J.; Moyer, N. Radon Pressure Differential Project, Phase I. FSEC-CR-344-90. Florida Solar Energy Center, Cape Canaveral, FL 32920, November 1990.
3. Cummings, J.B.; Moyer, N.; Tooley, J.J. Radon Pressure Differential Project, Phase II. FSEC-CR-370-90. Florida Solar Energy Center, Cape Canaveral, FL 32920, November 1990.
4. Energy Design Update, Cutter Information Corporation. 37 Broadway, Suite 1, Arlington, MA 02474-5552.
5. Florida Energy Extension Service. Insulation and the Building System—Build Green & Profit. Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, 2001.
6. *Home Energy Magazine*, 2124 Kittredge Street, No. 95, Berkeley, CA 94704. http://www.homeenergy.org.
7. *The Journal of Light Construction*. 186 Allen Brook Lane, Williston, VT 05495. http://www.jlconline.com
8. Lstiburek, J. *Builder's Guide (Hot-Humid Climates)* from the Building Science Corporation, 70 Main Street, Westford, MA 01886. 2000 (revised 2005), http://www.buildingscience.com/resources/books/default.htm
9. Lstiburek, J.; Carmody, J. (errata/update sheet 2002). *Moisture Control Handbook: Principles and Practices for Residential and Small Commercial Buildings*. Van Nostrand Reinhold Co., New York, NY, 1993. From the Building Science Corporation, 70 Main Street, Westford, MA 01886. http://www.buildingscience.com/resources/books/default.htm
10. Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831. http://search.ornl.gov.
11. Tiller, J.S.; Creech, D.B. *Builder's Guide to Energy Efficient Homes in Georgia*, 3rd Ed.; Georgia Environmental Facilities Authority, Division of Energy Resources: Atlanta, GA, 1999.
12. Tooley, J. The high-performance house—what does it take. Home Energy, March/April **1999**, *16* (2), 1. (Available online at: http://homeenergy.org/archive/hem.dis.anl.gov/eehem/99/990308.html).
13. MNL 18. *Manual on Moisture Control in Buildings*. American Society for Testing and Materials (ASTM). 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959. Treschel, H.Ed., 1994. May be ordered online from ASTM at: http://www.astm.org/cgi-bin/SoftCart.exe/BOOKSTORE/PUBS/30.htm?L+mystore+biad8244.
14. Volume 2—The Rehab Guide: Exterior Walls. HUD-DU100C000005956. U.S. Department of Housing and Urban Development, Washington, DC. August 1999. Available online at: http://www.huduser.org/publications/destech/walls.html.
15. *Energy Efficient Building Construction in Florida*. Florida Energy Extension Service, University of Florida, Gainesville, FL, Adapted with permission, 2005.

Facility Energy Efficiency and Controls: Automobile Technology Applications

Barney L. Capehart
Department of Industrial and Systems Engineering, University of Florida College of Engineering, Gainesville, Florida, U.S.A.

Lynne C. Capehart
Consultant and Technical Writing Specialist, Gainesville, Florida, U.S.A.

Abstract
This article examines the information and control technology used in new vehicles and points out the potential for using similar information and control technology in new buildings. The authors draw on their knowledge of new cars and new buildings to present a list of information and control functions together with the available sensors, computers, controls, and displays used in new cars that can provide significant opportunities for new buildings. Methods for integrating this new technology into new buildings are also discussed. The use of information and control technology in new cars should serve as a model for new building technology. This potential for new buildings should be recognized and similar technological improvements should be implemented.

INTRODUCTION

A great deal of new technology is available for buildings. The labels "Smart Buildings" and "Intelligent Buildings" have been around for years. Unfortunately, this wealth of new technology for buildings only exists in pieces and with products from many different companies—virtually no building constructed today utilizes a significant amount of this new technology. Most new buildings operate just like the buildings of the 1970s. Even though new materials, new design and construction methods, and new American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) building codes have greatly improved new buildings, these buildings still look and function much as they did 20 years ago. While most new buildings do have new equipment and better insulation, there is little in the way of new controls and display technology for the building occupants to see and use. Individuals seldom have the ability to control personal comfort and preferences.

In contrast, every new automobile—regardless of its price—is filled with new technology compared to the automobile of the 1970s. A new car typically comes with about 50 separate computers or microprocessors, has around 40–50 sensors, and provides about 20 electronic display and control functions. It does this for as little as $20,000. These automotive information and control systems commonly require little or no maintenance or repair for a period of three to five years. The technology is often visible, it can be used by the driver and passengers, it is generally standard on all new cars, and it is inexpensive and reliable. There is much fancier technology available if one want to pay for it (Lincoln Navigators, 7-Series BMWs, and S-Class Mercedes have around 100 processors onboard), but the majority of new automotive technology is found on every new car. The authors bought a new Toyota Prius Hybrid in early 2006, and it is like driving a computer! It even has a computer display console as a major feature of the dashboard.

With all of this new technology, today's cars are much more reliable and have significantly reduced maintenance requirements. In the 1970s, an automobile needed a tune up every 10,000 miles. Today, a typical new car does not need a tune up for 100,000 miles. Older cars needed new brakes about every 20,000 miles and now it's every 50,000 miles. The authors bought a new minivan in 1998 and did not have to take it back to the dealer for service for 40,000 miles! The vehicle had several oil changes in that period, but it needed no mechanical or electrical work.

In comparison, buildings need maintenance people from the moment they are first used. This does not include janitorial work, but the maintenance of lights, air conditioners, switches, controls, doors, and windows. This is like in the old days with cars—when people started making lists of things to be fixed as soon as they drove the car off the dealer's lot. People are paying extra for building commissioning just to make sure everything in the building is operating correctly and that it is fixed if it is not. Why can't a new building operate for six months, a year, or even several years without needing any maintenance? Cars do.

Keywords: Car technology; Controls; Displays; Building control systems; Dashboards; Smart sensors; Control modules; Individual controls; Automotive controls and displays.

What is the potential for using reliable, comprehensive, integrated, and inexpensive components in new buildings to create a transparent and efficient information and control system? What should be done in terms of buying new buildings? Clearly, progress in adapting and implementing technology for new buildings has a long way to go. Nonetheless, societies should demand more technology—a lot more. Technological improvements should be standard features that come with every new building without question rather than being options that add significant cost to the building. The only question should be where does one draw the line between standard features and additional new technology that will cost more?

FEATURES OF AUTOMOBILES THAT COULD BE USED IN BUILDINGS

Individual Control Systems

One of the most noticeable features of new automobile technology is how it provides the driver and often the passengers with individual control systems. Compared to a building, a new car has far more sensors, controls, and displays in a much smaller space. There are individually controllable air supplies for the driver and for the front passenger. Large vehicles often have air controls for the rear seat passengers too. Temperature ranges for heating or air conditioning are individually controllable, often for the front passenger as well as the driver. The air velocity is controllable with a multispeed fan. The outlet vents are easily reached and can be moved to direct the airflow onto or away from the person. The amount of outside air can be controlled by selecting fresh air or recirculation. Some lights, such as headlights and interior dome lights, are activated by sensors. Other lights are individually controllable. The driver or the passenger can turn on selected interior lights, can often dim these lights, and can direct the light to the area where it is needed. The moon roof can be opened or closed by the driver or the front passenger. Both front seats are individually adjustable for horizontal position, height, tilt, and back support, and many are heated as well. In addition, in some cars, these individual settings or preferences for functions like heating, ventilation and air conditioning system (HVAC) and seat positions are provided through a memory setting tied to an electronic key and settings for more than one person can be stored in memory.

Compare this technology to the control systems currently available in a common new building. A typical room in a new building may have a thermostat with a control set point and a temperature display at that location. It also usually has an unseen variable air volume system (VAV) control function and in a few instances a humidistat with a set point control and a display of the relative humidity at that location. Lighting is controlled with a single light switch or possibly by a single occupancy sensor. Otherwise, the occupants usually have no other sensors, controls, or displays in that room.

An example of a new technology that is currently available and that achieves some of the goals of individual control over personal space within a building comes from Johnson Controls. Their Personal Environments system is an easy to use desktop control unit that gives each person the flexibility to adjust temperature, lighting, air flow, and acoustic characteristics as often as necessary to maintain personal comfort levels. Individuals can adjust the air temperature and control the amount and direction of air flow at their desktop. They have a heating panel under the desk to adjust the temperature to their legs and feet. The Personal Environments system also allows an individual to control task lighting and to mask background noise. The system has a sensor that turns off all functions when the workstation is unoccupied for more than 10–15 minutes. Although this system is being used by a number of companies, it is the exception rather than the rule.

Operational Controls

In addition to personal comfort controls, the new car also has a large number of automatic control systems to optimize and control its own operation. Engine control systems insure fuel efficiency and reduce air pollutants from the combustion process. Sensors for inlet air temperature and relative humidity allow optimum fuel flow control and optimum combustion. System computer modules also control the anti-lock braking system (ABS), transmission, cruise control, and body controller. These microprocessor systems are standard on new vehicles, but new buildings are not built the same way. Operational controls are available for new buildings but they require special design criteria. No one considers the possibility that they should be standard equipment.

Display Systems

New cars tell the owner about much of the maintenance and repair that needs to be done and they certainly notify the driver whenever one of the major systems is in need of attention. A new car has sensors that report tire pressure, unclosed doors, lights or other controls left on, unfastened seatbelts, brake fluid status, and many other operational features related to the safety of the car and the occupants. Most new cars are also giving drivers and occupants information about the operation of the car, including real time fuel economy as well as average fuel economy over a selected mileage. Even a cursory comparison shows that new buildings lag very far behind the present use of technology in new cars.

Much of the information on car maintenance and safety is aimed at the driver. What comparable information does a building operator get about the maintenance needs of the building or the various rooms in a building? Things that would be helpful to know include whether the air handling system filters are dirty, whether the refrigerant is at the proper level, whether sensors are working properly, whether lights are burned out, or whether the doors have been left open.

The present system in buildings is essentially a manual system. Filters are checked by maintenance personnel on a time schedule. Maintenance workers often depend on "human" sensors to notify them of burned out lights, improperly functioning photo sensors, or temperature problems in individual rooms.

Options

New cars have options and new buildings have options, but these mean very different things. An option for a new car is an item or function that is already available and can be installed on the car at extra cost. For a building, an option is an item or function that an owner wants to add at extra cost, but expensive additional design, engineering integration, and testing work must usually be performed before it can be installed and operated.

HOW DID THE AUTOMOTIVE INDUSTRY DO THIS?

It is important to understand how new automobiles can have so much new technology at such a low cost and why they are so reliable in order to know how to utilize similar innovations in the building industry.

Engineering Analysis and Design

A significant amount of engineering analysis and design goes into both the structural and operational features of a new car. In addition, significant engineering analysis and design also go into the manufacturing and production processes for assembling the new cars. A major benefit of this approach is that the car's entire system and subsystems, as well as each of the car's components, are carefully engineered. For example, the electrical power consumption of the components and systems in a new car are carefully analyzed—built and selected to make sure that the total power demand is not greater than the capacity of the electrical power supply system, i.e., the 12-volt battery. Thus, with cars, the need for energy efficient electrical systems is built in from the start.

When a building is designed, the electrical load is specified first, and then a power supply system that is big enough to handle the load of the building is specified. Little or no thought is given to minimizing the electrical load itself because there are generally no constraints on the amount of power a utility will supply to the building.

Overall Quality Control Programs

A new car is reliable because a significant amount of engineering goes into both the car design and its manufacturing process. Quality and quality control start with the engineering design and are strongly emphasized throughout the manufacturing and assembly of the car. Individual components are designed and made with quality and reliability as major goals. Subsystems and final systems—including the entire car—are similarly produced. Ordinary and accelerated life testing are conducted on the car's components, subsystems, and systems. These extensive tests include the effects of temperature, moisture, mechanical and thermal stress, and other factors. As a result, most of the car's components and systems will last at least three years or for 36,000 miles. Warranties on some new cars are now available for seven years or 70,000 miles.

Quality control and warranties in building design and construction are very different. The auto manufacturer provides the warranty for the entire vehicle (with the possible exception of the tires); the systems in new buildings are likely to be under several different warranties. The HVAC manufacturer covers the HVAC system, the flooring manufacturer guarantees the carpet/flooring, the plumbing manufacturer guarantees the plumbing fixtures, etc. There is usually no centralized quality control or warranty for a new building as with cars.

Widespread Use of Microprocessors and Computers

Much of the technology and operational features of new cars comes from the use of microprocessors and microcomputers. A new car may have as many as 50 separate microprocessors and 11 major computer-based systems. Some new luxury cars have up to 90 microprocessors. It is often said that a new car has more computer power in it than the first manned space capsule did. Computer-based systems are found in the system modules for new cars and they account for much of the engine performance, reduced emissions, sophisticated diagnostics, and many comfort and convenience features. The Engine Control Unit (ECU) is the most powerful computer in the car and it has the demanding job of controlling fuel economy, controlling emissions from the engine and the catalytic converter, and determining optimum ignition timing and fuel injection parameters. These computers, microprocessors, and system modules greatly simplify the diagnostic job of finding problems with the car and provide information on what kind of repair or replacement work is needed.

While a large new building with a sophisticated building automation system (BAS) may well contain 50 or more microprocessors, this does not match the new car in terms of having equal computing power per room or per group of rooms with two to four occupants. The rooms and offices in buildings do not have monitoring and self-diagnostic features. They could—the technology, equipment, and systems exist—but they are not supplied as a standard item and they are not available in the same way that options available on new cars are.

System Modules

As discussed above, the system modules are where the computer-based systems reside in new cars. These system modules are highly complex and highly important systems in new cars. Many highly desirable performance and comfort features are provided by system modules. Typical system modules in a new car are the Engine Control Unit, the instrument panel module, the climate control module, the transmission control module, the power distribution box module, the airbag module, the driver's door module, the ABS module, the body controller module, and the cruise control module. These are the system modules found on every basic car. Additional system modules are options for lower priced cars or standard features for higher priced cars. These include navigation control modules, entertainment system modules, advanced comfort control modules, and communication control modules for computers, cell phones, and Internet access.

Communications Buses

Using standardized communications buses with these system modules makes both designing and building new cars much easier than it was in the old days. Two major services must be accessible to every area of a new car—electric power and the communications bus. All of a car's system modules must be able to communicate with each other, receive signals from most of the sensors in the car, and send signals to the control components, systems, and actuators. Using a communications bus greatly simplifies the wiring, reduces the number of data sensors, and implements additional features at very little additional cost. Without the communications bus, the job of wiring up a car during the assembly operation would simply be too labor intensive and time consuming for a reasonable-cost product. Also, the speed of communications is so important now that only a digital bus has the speed and capacity to handle the data collection and data transfer load for a new car.

The communications bus and the system modules work together to make the design and building of the car much easier. Information is sent over the communications bus in a standard communications protocol—usually the SAE J1850 standard or the Controller-Area Network (CAN) standard, although some manufacturers are using FlexRay, which is a faster and more sophisticated communications bus. Data is sent in packets with a standard structure—a label and some data. For example, an information packet with Speed for the label and 52.5 for the speed data in MPH is picked up by the instrument control module, which refreshes the indication on the speedometer with this new data. The standard communications bus makes the design of the various system modules much more straightforward. In addition, the sensors in the car only need to send packets of data to the communications bus; therefore, the car maker does not have to deal with the problem of a particular sensor putting out a strange voltage or a current signal that must be converted somewhere into a true physical parameter of the car's operation. In this example, the alternative is to tell the instrument panel module maker that the signal for speed was going to be a 4–20 mA current loop value and that 10 mA was equivalent to 40 MPH.

The use of the standardized communications bus also makes it easy to use outside suppliers and sources for many of the components and systems in a new car. The carmakers do not have to worry about how a specific sensor or module works internally, they only need to know that the data will be transmitted in a known, standardized manner, and that it will have a known, standardized structure. Much of the success with using modern technology in cars and much of the reliability of that technology comes from using the simplified approach of a standardized communications bus.

This same type of technology is essentially available for new buildings. BACnet, LONWorks, and TCP/IP are the most common standard communication protocols. TCP/IP may be the ultimate answer, but another level of standardization is also needed to insure the data that comes across TCP/IP means the same thing to each different piece of equipment in a facility. Most buildings are being wired for a Local Area Network (LAN) using either coaxial cable or fiber optic cable. Thus, the hardware and software are available, but there is no organization responsible for requiring or enforcing the standardized interconnection of all of the building components, subsystems, and systems, like the automakers have. Without a standardized communications bus running through the entire facility, together with accessible electric power, buildings will never have the kind of technology that cars have and will never have the cost benefit or the reliability that this kind of technology can bring to buildings.

Smart Sensors

Most of the basic automobile sensors that were used in the past to read continuous physical parameters such as temperatures, pressures, flows, and levels operated on the principle of producing a voltage or current output proportional to the real value of the parameter. The output of these sensors was almost always nonlinear and also varied with the temperature or other physical parameters. This resulted in poor measurements or required the use of more equipment and processing power to correct the sensor reading for the nonlinearity and to provide temperature compensation curves to get accurate readings. Today, smart sensors are used to provide these functions and to output data to a microprocessor or system module. The sensor output is input to the microprocessor and the sensor reading is digitized, corrected, temperature compensated, and sent out over the standardized communications bus.

These smart sensors interface directly to the communications bus and provide fast and accurate measurements. Because the sensor package contains a microprocessor, much of the load is removed from the system module that the smart sensor is supporting. Designed and built as an integrated package, the smart sensor fulfills its mission reliably with a low initial cost.

The sensors for buildings are expensive and many of them are not very reliable. They are certainly not reliable in comparison to sensors in cars. In particular, the relative humidity (RH) sensors and CO_2 sensors are notoriously unreliable and require frequent cleaning, calibration, and general maintenance. That level of performance would be unacceptable for these sensors in a car. Why shouldn't the sensors in buildings work reliably for a period of three to five years before they need any significant attention?

Wiring Harnesses and Standard Connectors

The use of preassembled wiring harnesses and standard connectors has made the task of wiring up a new car much easier. It is important to use stranded, not solid, wire cable. Each length of stranded wire consists of a twisted bundle of very thin thread-like wires. Solid wire, on the other hand, is a single thick wire segment. The advantage of stranded wire is that it is much more flexible than solid wire and it also less susceptible to breakage. One thread of a stranded wire can break without affecting the performance of the connection, but if a solid wire breaks the connection is lost. Also, if there is one weak link in the reliable performance of any electrical or electronic system, it is the connectors. With this in mind, the importance of carefully and correctly built standardized connectors cannot be overemphasized.

Use of Skilled Assembly Workers

The auto industry has a large supply of skilled workers for its design, engineering, and assembly operations. These skilled workers receive training in their specific jobs as well as training in quality control and process improvement techniques. Many of the manufacturing and design improvements in new cars have come from the production workers themselves. In addition, skilled workers have made a great improvement in the overall reliability and quality of the new cars. Autoworkers are usually paid more than those working in other industries or services.

Problems with the construction of new buildings often come from the use of workers with minimal or insufficient skills for the job. Finding skilled workers may be difficult, and is certainly expensive. The nature of building construction often impedes the retention of skilled workers. As a result, there may not be a large pool of highly qualified building construction workers available when a particular building is being built.

One of the most common problems in building structures is the roofing, which is the subject of the greatest number of lawsuits in building construction. Most roofs leak, and they leak from the day the building is first occupied. Roof leaks are the result of poor installation and construction rather than problems with roofing technology and materials. When a roof leaks, water leaks into the walls and may not be noticed until mildew and rot are visible. By then, the building may be significantly damaged. Mold, mildew, and indoor air quality (IAQ) problems in the building will require more time and money to fix. Using sensors in new buildings to identify roof and wall leaks when they occur is a critical application of automotive-type technology in new buildings. New cars use infrared reflectance sensors to identify rainfall on windshields and they automatically start up the windshield wipers. These sensors, or other types of moisture sensors, if installed throughout new buildings, would quickly identify leaks and moisture buildup and alert building operational people to this serious problem.

Poor workmanship can cause many other problems in buildings. Even the HVAC system can be affected because random testing has shown that many air conditioning systems are installed with an improper charge of refrigerant. In economic terms, the problem of workers with insufficient skills and workers who have had poor quality training results in the need to commission buildings to check and see if the building components and systems work as they should. (See the discussion on Commissioning below.) This expense is clearly attributable to a lack of adequate engineering, a lack of quality control measures, and especially a lack of highly trained workers.

WHY DOESN'T NEW BUILDING CONSTRUCTION INCLUDE MORE NEW TECHNOLOGY AS STANDARD EQUIPMENT AND SYSTEMS?

Automobiles are built according to a standard plan; building architects, on the other hand, reinvent the wheel each time they design another building. This lack of standardization in buildings impedes the introduction of new technology in new building construction. Other factors also influence this difference in approach.

Unlike new cars, most new buildings are site-built, and are built to "cookie cutter" specifications that emphasize the lowest first cost of construction. Even "custom built" buildings are held hostage to the lowest first cost syndrome. Thousands of different construction companies build residential and commercial buildings. Hundreds of different companies build fairly large commercial buildings. These companies range in size from small businesses to major architectural and engineering firms and major construction firms. It is extremely difficult to implement standards of technology when this many individual companies are involved.

The fact that most buildings are site-built impedes the assembly line and systems approach to installing new technology used in the auto business. One area of building construction that is immediately amenable to the assembly line approach of the carmakers is the construction of prefabricated or modular buildings. This manufacturing sector could easily incorporate the knowledge from the automotive assembly sector to produce buildings with the same level of technology and reliability as new cars. The engineering and quality control functions are much more cost effective in this sector. This sector could easily use more computers, more microprocessors, more system modules, more smart sensors, and a standardized communications bus.

Cars are Constructed in a Factory Assembly Line and Moved to Their Ultimate Market and User

The factory environment makes it easier to train workers to install the equipment in new cars as well as training them in quality control procedures. Buildings, however, are constructed at the point of use. Construction workers may work for a long time on a single building doing all types of work. Their training is not likely to be technology specific. Auto assembly workers typically specialize in some part of the assembly process and therefore can be trained on this more limited work task. In addition, they become quite knowledgeable on this part of the assembly operation and soon become able to add value to the company by suggesting improved methods of designing and constructing components and systems that they assemble. Quality control is more easily stressed in this environment and many of the workers actually see the final result of their work drive off the assembly line, which serves to positively reinforce the need for a high skill level and the need to perform high quality work. In fact, these workers often own and drive cars produced by the company they work for. They are more likely to reject poor quality parts, components, systems, and assembly procedures.

More new cars are sold each year than new buildings, so there is a larger market for the technology and the price can be reduced due to bulk purchase of the equipment. This is certainly true at face value, but when the scale of use of technology for buildings is considered, the numerical superiority of the cars goes away. If one considers that the unit of interest in buildings is rooms and that the interest is in having the same technology level in each room that is in cars, the perspective is now very different. There may very well be more rooms than cars built each year. Thus, the comparison of a room to the car rather than a building to a car will lead to a much greater economy of scale for new building construction and should provide a strong economic incentive to move in this direction for buildings.

Cars have a shorter lifetime than buildings, so new technology can be introduced faster and the customers can develop a faster appreciation for what it does. Cars do have a shorter lifetime than buildings, but most buildings end up being refurbished or having the equipment and systems retrofitted, so there is still a lot of opportunity to use new technology in older buildings. Sensors, controls, system modules, and many of the other features of new car technology can be added to older buildings when they are needed. In general, the most cost effective way to build an energy-efficient and functionally superior building is to do it right the first time rather than retrofit it later. However, new equipment (especially new information technology) can be added to rooms and to the entire building. It would have been easier and cheaper to install coaxial or fiber optic cable in a building when it was built, but people still have managed to find a way to get the LAN cable and connections into rooms and offices so PCs can be networked.

Purchasers of new cars are influenced by features they have seen on other cars. Therefore, consumer demand is important in increasing the marketability of new technology options. This is one reason it is necessary to start installing some of this new technology in buildings. Once building owners, managers, and occupants start seeing what has been done in other buildings and how much more enjoyable and productive it is to work in buildings with this advanced technology, they will start to demand more technology as a result. It is somewhat amazing that the people who drive cars with all this new technology will go happily to work in buildings that do not come close to providing similar comfort and operational features of automobile technology!

Cars are Designed for Use by Individuals; Buildings are Designed for Use by Companies

The motivation of the designers and the manufacturers of cars is frequently different from that of people who design and build buildings. Car manufacturers build a car to attract a buyer—they add bells and whistles to make their car different. They encourage innovation and thinking outside the box. Architects and construction companies are building a box, so their thinking often stays in the box. They may work on the exterior design or they may make the interior appearance pleasing, but they do not think very hard about what goes on inside the box and they don’t consider the needs of the individuals living and working in the box. A car designer should consider safety when drawing up plans for a new car. Beyond putting in emergency exits and sprinkler systems, a building designer may not think about how to make the building safer because that is not part of the job. Among the questions that building designers should be asking are: "How can this building be made more comfortable, safer, and more user-friendly?" "How can occupants interact with this building to increase their comfort and safety levels?" "How can we make this a building of the future?" With a little imagination and an increased use of information and controls technology, building designers can make significant changes in the comfort level of the occupants.

WHAT DOES THE BUILDING CONSTRUCTION INDUSTRY NEED TO DO?

Establish an Integrated Design-and-Build Engineering and Management Structure

The amount of engineering work that goes into a new building must increase significantly. The building structure should be designed with high technology use in mind and it should utilize new technology to deliver the performance and comfort features that are desired in new buildings. In addition, quality control and reliability should be designed and engineered into the building from the start of the project. Then, quality management techniques should be employed so that the building is actually constructed to provide the quality and reliability features that are expected.

Use Equipment and System Modules in New Buildings

This approach has facilitated the use of most new technology in new cars at a reasonable cost and with extremely good reliability. However, the standardized communications bus has made the most dramatic difference. By using a standardized communications bus and system modules, car technology could be transferred to buildings relatively easily. Individual HVAC modules for occupants, individual lighting modules, other comfort modules (such as for seating), and building operation and maintenance modules could all be used to greatly increase the performance and reliability of new buildings and yet allow them to be built at reasonable costs. Certain sectors such as the residential manufactured housing sector and the hotel/motel sector as well as many office buildings could easily adopt this approach.

Even site-built homes could incorporate some of these features. Residences are often prewired for intercoms, telephones, security systems, cable, and high-speed internet connections. Designing a central integrated system for monitoring and controlling the performance and comfort of a home and prewiring the house for such a system is well within the realm of feasibility. It is possible to envision a home with a central control panel that is accessible from the Internet. Homeowners could monitor their homes from work. They could receive security or fire alarms. They could make changes to thermostat settings if they knew they were going to be early or late getting home. They could receive alarms if there was a water leak or if the refrigerator stopped running.

Build More Modular Buildings

The solutions for providing greater use of technology in new buildings and for providing quality and reliable buildings are much easier for modular buildings with significant presite construction performed in a factory or controlled environment. High-tech components and equipment can be installed more easily in prefabricated and modular buildings within a controlled environment with highly skilled and quality control trained workers.

Impose Standards on Equipment and System Suppliers

Most major construction companies are already in a position to do this—they have the financial leverage to specify the components and equipment that meet their exact requirements. The residential manufactured housing sector in particular could do this quite easily. The federal sector, states, and large companies also have excellent opportunities to set these standards. One of the most important standards is to require a standardized communications bus in a building with all sensors and controls interfacing directly with that communications bus.

Support Codes, Standards, or Legislation to Increase the Use of New Technology in Buildings

Building codes and standards have been responsible for many of the improvements in standard buildings. With minimum equipment efficiencies, minimum thermal transfer levels, and minimum structural standards in place, companies that construct buildings must meet these minimum standards regardless of whether it increases the first cost of the building. Without minimum standards such as the ASHRAE 90.1 standard, many buildings would still have inferior equipment and poor insulation because it was cheaper to put in initially. Other programs like Leadership in Energy and Environmental Design (LEED) and EnergyStar could incorporate requirements for adding new comfort and control technology in buildings. The standards for utilizing new technology could be set voluntarily by large companies and big purchasers of buildings like the federal sector, states, schools, and the hotel/ motel sector. The auto industry has certainly incorporated many of the new technological features without needing government intervention.

Integrate New Building Technology with the Desktop Computers and Building Automation System that are Already Being Installed in New Buildings

The types of smart sensors, system modules, and standardized communications buses that the authors have been recommending for use in new buildings should be considered an integral part of the overall Building Automation System. All of the components, systems, and equipment must work together seamlessly to provide the expected level of performance and comfort and all the desktop computers should be tied into these systems through a Local Area Network.

The desktop computer could be the equivalent of the car dashboard or instrument panel, and it should be the personal interface to an expanded building automation system (BAS). It could tell what the space temperature is and how much ventilation is being provided. It should allow occupants to set their personal preferences for lighting levels, seat positions, window or skylight openings, etc. It should also let them enter new desired values of these space parameters.

BENEFITS OF STANDARDIZED COMMISSIONING OF BUILDINGS

Commissioning a building is defined in ASHRAE Guideline 1 (1996) as the processes of ensuring that building systems are designed, installed, functionally tested over a full range, and capable of being operated and maintained to perform in conformity with the design intent (meaning the design requirements of the building). Commissioning starts with planning and includes design, construction, start-up, acceptance, and training, and it can be applied throughout the life of the building.

Commissioning a building involves inspection, testing, measurement, and verification of all building functions and operations. It is expensive and time consuming, but it is necessary to insure that all building systems and functions operate according to the original design intent of the building. Commissioning studies on new buildings routinely find problems such as control switches wired backwards, valves installed backwards, control set points incorrectly entered, time schedules entered incorrectly, bypass valves permanently open, ventilation fans wired permanently on, simultaneous heating and cooling occurring, building pressurization actually negative, incorrect lighting ballasts installed, pumps running backwards, variable speed drives bypassed, hot and cold water lines connected backwards, and control dampers permanently fully open. And this is only a short list!

The process of commissioning a building constructed like a new car and using the new car-type technology would be far quicker and simpler as well as much less expensive. The use of standardized components, subsystems, and systems could actually eliminate the need to check and test these items each time they are used in a new building. In a factory or laboratory, standardized commissioning tests could well determine their acceptability with a one-time procedure. The use of a standardized communications bus would dramatically shorten the time and effort of on-site testing of the building components, subsystems, and systems. Data from all sensors and controls would be accessible on the communications bus and would allow a significant amount of automated testing of basic functions and complex control actions and responses in the building. A commissioning module could also be added to the building systems, and this would even further automate and speed up the commissioning process. This commissioning module would remain as a permanent building system and it would not only aid in the initial commissioning process but also the recommissioning process and the continuous commissioning process.

Presently, the cost of commissioning a new building is around two to five percent of the original cost of construction. The use of standardized commissioning tests and the use of a commissioning module would greatly reduce this cost. Although commissioning is a cost effective process—usually having a payback time of one to two years—many building owners do not want to spend the additional money for the commissioning effort. A prevailing attitude is "I have already paid to have the job done correctly. Why should I have to be the one to pay to check to see that it has actually been done correctly?" This is a difficult attitude to overcome, and it is often a hard sell to convince new building owners that they will actually come out ahead by paying to verify that their building does work as it was designed to work.

One final note on commissioning is from one of the author's energy audit experience. Many problems found when conducting audits of existing buildings are clearly problems that have been there since the building was constructed. For example, in the audit of a newspaper publishing company, it was found that the cost of air conditioning was excessive. Further checking showed that the heating coil and the cooling coil of the major air handling unit were both active during the hottest part of the summer. The control specifications specifically called for simultaneous heating and then cooling! Once that original problem was corrected, not only did the air conditioning bill go down dramatically, but the building occupants reported that they thought the air conditioning system was working much better because they were much more comfortable.

DO NEW BUILDINGS NEED "DASHBOARDS"?

The dashboard and instrument panel is the heart of the driver-car interface. Status information on the car's operation is presented there in easily understood form. A similar feature in a new building would make sense—not the complex human-machine interface (HMI) or graphical user interface (GUI) from a BAS, but a simplified display for average building occupants and maybe even one for the building engineer or maintenance supervisor. Each floor of a building could have a "dashboard" display. It could be located in a visible place and occupants could see the status of energy use in terms of peak cost or off-peak cost, daily use of kWh, and therms of gas. They could also see temperature and RH conditions at a glance and they could get red light/green light indicators for energy use and maintenance actions. Several of these "dashboards" could be provided to the operation and maintenance staff. These simplified "dashboard" displays could also be available on the PCs of the occupants and operating personnel. Cars provide a powerful model to use in many building technology applications.

CONCLUSION

New buildings have not kept up with technological advances, especially when compared to automobiles. All one needs to do is make one trip in a typical new car and then make one visit to a typical new building to see this themselves. Comfort levels, safety levels, reliability levels, quality control levels, and automation levels are all much higher in new cars than in buildings. The imagination and creativity that goes into new car technology and manufacture should be harnessed for new buildings as well. It is necessary to start building new buildings in the same way new cars are built.

ACKNOWLEDGMENTS

Much of the material in this entry was taken from a paper titled "If Buildings Were Built Like Cars—The Potential for Information and Control System Technology in New Buildings," by Barney L. Capehart, Harry Indig, and Lynne C. Capehart, *Strategic Planning for Energy and the Environment*, Fall 2004, Fairmont Press; and "Why Can't A Building Think Like a Car?—Information and Control Systems Opportunities in New Buildings," Chapter 38 in Web-Based Energy Information and Control Systems, Edited by Barney L Capehart and Lynne C Capehart, 2005, Fairmont Press. Both used with permission from Fairmont Press.

BIBLIOGRAPHY

1. Argonne National Laboratory program on sensor development, http://www.transportation.anl.gov/ttrdc/sensors/gas-sensors.html.
2. Automated Buildings, http://www.automatedbuildings.com.
3. Court Manager's Information Display System, http://www.ncsc.dni.us/bulletin/V09n01.htm.
4. Delphi Automotive Electronics, http://www.delphi.com/automotive/electronics.
5. How Car Computers Work, http://www.howstuffworks.com/car-computer.htm.
6. Motoring with Microprocessors, by Jim Turley, http://www.embedded.com/showArticle.jhtml?articleID=13000166.
7. New car features, http://www.autoweb.com.au.
8. Sensors, Automotive Engineering International Online, http://www.sae.org/automag/sensors.
9. Smart Energy Distribution and Consumption in Buildings, CITRIS—Center for Information Technology Research in the Interest of Society, http://www.citris.berkeley.edu/SmartEnergy/SmartEnergy.html.
10. Today's Automobile: A Computer on Wheels, Alliance of Automobile Manufacturers, http://www.autoalliance.org/archives/000131.html.

Facility Energy Use: Analysis☆

Klaus-Dieter E. Pawlik
Accenture, St Petersburg, Florida, U.S.A.
Lynne C. Capehart
Consultant and Technical Writing Specialist, Gainesville, Florida, U.S.A.
Barney L. Capehart
Department of Industrial and Systems Engineering, University of Florida College of Engineering, Gainesville, Florida, U.S.A.

Abstract

When you do an energy audit, you must have a good understanding of how the energy is used by a facility to make sure that your energy efficiency recommendations are accurate and appropriate. You must know what equipment uses energy, how much energy it uses and how much energy it uses as a proportion of the total energy used at the facility. You can do this by monitoring the energy use of all the major pieces of equipment, but this is quite expensive and time-consuming. In this article, Part I, we discuss an alternative way for you to estimate the energy use by developing an "Energy Balance." This method is particularly helpful if you have limited time to gather the energy use data.

At the University of Florida Industrial Assessment Center (UF IAC) we devised the energy balance method of energy analysis because we realized that we had sometimes overestimated the energy savings available to a facility. This happened when we did not have a clear idea of how much energy was used by each group of equipment such as motors, lighting, air compressors, air conditioning, etc. Therefore, we developed a spreadsheet where we entered the energy use data gathered at the facility and compared it to the actual energy use as shown in the utility bills.

We then refined our assumptions with respect to load factors and hours of usage if our preliminary results were out of line with the actual total energy use by the facility. Our result is an approximation of the energy used by the major electric equipment systems in the facility. We perform separate energy balances for each type of energy source (electric, natural gas, propane, etc.).

Performing an energy balance gives you several benefits. You can see which equipment systems contribute the most to the energy costs and can spend your time analyzing recommendations for those systems. Since the equipment data is already available in the energy balance spreadsheet, you can easily export portions of the data to the recommendation analysis spreadsheets to calculate energy cost savings. You can check the credibility of your energy savings recommendations by making sure that the recommendation does not save more energy than a given equipment system uses.

Finally, you can show the facility how its energy use and energy dollars are apportioned among the various systems.

In this article, we describe the spreadsheet we developed to estimate an energy balance along with several sample energy balances. We explain how to reconcile the preliminary results to achieve a reasonably realistic energy balance. We also provide examples of recommendations that we had to reanalyze and basic assumptions that we needed to revise for consistency with the energy balance.

INTRODUCTION

To improve the efficiency of energy use at a facility, the analyst must have a good understanding of what that energy use is. Energy use data can be gathered by metering and monitoring equipment, but this method is expensive and time-consuming. Often, the energy analyst can spend only a day gathering data, and this sometimes makes getting an accurate picture of the facility's energy use difficult.

At the UF IAC, we were faced with the task of gathering sufficient data in a 1-day plant visit, to analyze the existing energy use and make recommendations for changes. We developed an energy balance approach to simulating facility energy use to help with our analysis.

This article will demonstrate the usefulness of constructing an energy balance as an alternative to more

☆ This entry originally appeared as "Analyzing Facility Energy Use: A Balancing Act" in *Strategic Planning for Energy and the Environment*, Vol. 21, No. 2, 2001. Reprinted with permission from AEE/Fairmont Press.

Keywords: Energy use analysis; Energy balance; Estimation of energy use; Energy data analysis; Energy accounting; Load factors; Use factors; Equipment ratings.

costly methods of measuring energy use. We will detail the major steps of creating and using an energy balance. We start with data collection, explain the method used to reconcile the simulated use with the historical use patterns, and discuss use of the energy balance for making recommendations.

WHAT IS AN ENERGY BALANCE?

We modeled our energy balance after the traditional mass and heat balance approach of physics and thermodynamics. We realized that it was not sufficient to look at each energy use system at a facility in isolation because we sometimes found situations where the whole did not equal the sum of its parts. That is, the total metered kWh (and/or kW) use was lower (or higher) than the sum of the estimated uses by each system within the facility. Therefore, we developed a model for estimating the energy uses of each system as a part of the whole system, and we call our model the energy balance.[1]

CONSTRUCTING A DETAILED ENERGY BALANCE SPREADSHEET

Before visiting the facility site, you must gather the historical energy use data either from the facility or from the utility supplying the energy. This data must be collected both to prepare for an audit and to start understanding the energy use of a facility. At the facility site visit, you should collect data on all energy-consuming equipment (electricity, gas, or any other form of energy). The equipment-energy use data should be entered in the energy balance spreadsheet. Next, you calculate demand and energy use for the facility based on that data, and adjust the equipment energy use parameters until the predicted use matches the historical use within a given tolerance factor. Only then should you start the data analysis, necessary to make energy efficiency recommendations.

Collecting Historical Energy Use Data

To analyze a facility's historical energy use, gather atleast 12 months of energy bills. For each type of energy source, find the average demand and energy uses and charges for the past year or business cycle. Do not forget to factor in the various taxes and tariffs. For electricity, demand charges are based on the peak recorded each month, while the energy charge is typically based on the number of kilowatt-hours used each month.

At the UF IAC, we used an excel spreadsheet for organizing monthly use and costs. Once the data has been entered, you can quickly produce useful bar graphs of demand and energy use as well as costs trended by the month. These graphs give a quick picture of historical energy use.

For example, you may find a bi-level peak pattern that may indicate a business with different seasonal usages, or you may see a continually increasing pattern. This type of picture will help the auditor understand the processes and business of the client and may prompt questions that plant personnel can answer during the plant visit.

For example, can a bi-level pattern of demand and energy be explained by different operating schedules during different seasons? Or, why is the use continually increasing? Perhaps increased production accounts for the increase, but maybe it is poor maintenance practices and aging equipment.

In addition to the bar graphs, we used pie charts to illustrate the relative energy use between different energy sources. For example, if a plant uses both natural gas and electric energy, then a pie chart makes it easy to see the relative usage between gas and electricity. If the analysis shows insignificant gas use (less than 10% or so), then you know that you should focus your major analytical efforts on the electrical equipment. You will also use this historical use data with your energy balance calculations in order to validate your data collection and assumptions.

Collecting Equipment Energy Use Data

We utilized the following basic method to collect data:

- Meet with plant personnel, someone like the plant manager, to discuss the processes at the plant as well as major energy-consuming equipment and safety concerns. We placed an emphasis on the plant process, since these processes are how the plant makes money. Any energy savings recommendation that causes the plant to lose money in other areas is sure to be rejected.
- Conduct a plant walk-through with plant personnel, again focusing on processes with the emphasis on major energy-consuming equipment and safety concerns. We generally followed the path that the raw material would take as the plant transforms it to a completed product.
- Perform a detailed data collection of energy-consuming equipment. If plant maintenance personnel will help with this task, their services are invaluable.

You will need the following basic data from all types of energy-consuming equipment to compile the energy balance:

- From the discussion with plant personnel, climate conditions, and personal observation, you can estimate the annual use in hours per year (h/yr) and assign a utilization factor for each piece of equipment. We chose the number of hours a year that a piece of equipment should be available for operation as the annual use number. For example, we often used the operating hours for the facility for the annual use hours. Then we modified this annual use number with a utilization factor (uf) to account for the actual time of use of that equipment. For equipment in constant use, the utilization factor could be 0.9 to account for 10% downtime due to maintenance. The utilization factor for equipment in less than constant use is assigned a value according to the percent of use time estimated by plant personnel. This parameter can be adjusted, if necessary, when refining the energy balance.

- From the name plate of the equipment, you should collect and record the following data if available:
  - The power rating in kilowatts (kW), horsepower (hp), tons, or similar information. For example, if neither kW nor hp is available, collect the amperage, voltage, power type, and power factor. By power type, we mean alternating current (a.c.) single-phase or three-phase, or direct current (d.c.) With this additional information, you can estimate the power rating.
  - Any efficiency values that may be available. For cooling equipment, you should record the energy efficiency ratio (EER) or coefficient of performance (COP). Motor efficiency values can be found on the nameplate or from tables in the DOE MotorMaster program.[2]
  - If no data are available, try to get the manufacturer's name, model and serial numbers, and contact the manufacturer for data.
- Since a motor rarely runs at its rated power level, you need to determine the motor load factor (lf) to calculate the actual power used. You can take a current measurement or a speed measurement of a major motor to get an approximate load factor. You could use a clamp-on ammeter or a tachometer and compare the measurement to the amperage rating or the speed drop characteristics curve to estimate its load factor. However, you must remember even this type of measurement is just a one time sample. The only way to get an accurate measure would be to measure load over a relatively long period. To get this accurate measurement, you could use a strip chart recorder, a monitor connected to a computer, a data logger, or part of an energy management system. If you have not measured the load factor, you will have to estimate one. We have found that typical motor load factors are between 0.35 and 0.60. However, some motors such as fans and air compressors generally have higher load factors of 0.8–0.9.
- To adjust the total power use, we also assign a diversity factor (df) to each piece of equipment. We use this to account for similar pieces of equipment that are not run concurrently. For example, if the plant has two condensate pumps but only one can be run at any given time and the second is used in a backup mode, you should account for this fact with a diversity factor of 0.5. Most equipment that operates a large number of hours during the plant's peak use shift will have a diversity factor of 1. (For a more complete explanation of load factor and diversity factor, see Capehart[1,3]).

It is not feasible to collect data on every piece of equipment at a facility. However, you should try to collect detailed information on the larger pieces of equipment that use 90% of the energy. We used a miscellaneous category to account for the last 10% of energy use by the large number of small energy users.

You should collect data on the following pieces of equipment:

- *Motors*. These are the largest single energy consumer category in most manufacturing facilities. Generally, you should concentrate on motors over 5 hp. However, depending on the size and type of facility, you may want to increase or decrease this threshold. Furthermore, you may want to treat a large number of similar small-size motors as a group. For example, some facilities may have a large number of conveyor belt motors of similar size and with similar usage patterns. See the motor section of Fig. 1: Energy Balance—Electrical Equipment.
- *Heating, ventilating, and air conditioning* (*HVAC*). Locate all the heating, cooling and ventilating equipment. Exhaust fans should be included in the motor list. Sizes of equipment (in kW, tons or other units) together with EERs or COPs must be found and recorded. From this EER or COP data, we can find the kW/ton power consumption and the total kW by knowing the size in tons of each unit. The energy consumption is then found by multiplying the full-load kW by a factor called the Full Load Equivalent Operating Hours (FLEOH) of the compressor unit. The FLEOH is dependent on the geographical area where the facility is located, and can usually be found from the local utility or from the State Energy Office for the area. (Note: you need to use the compressor operating hours, not the hours that the facility is air-conditioned, for your air conditioning equipment operating hours.) See the HVAC section of Fig. 1: Energy Balance—Electrical Equipment.
- *Air Compressors*. We placed air compressors in a separate category from other motors because the efficiency recommendations are often quite different and we wanted to isolate the air compressor system's energy use. The hp size of each compressor should be recorded. When you examine the air compressor system, also look for auxiliary equipment, such as air dryers. See the air compressor section of Fig. 1: Energy Balance—Electrical Equipment.
- *Lights*. Note the usage patterns of the lights as well as the usage patterns of the lighted areas. Then find the total wattage of the facility's lighting system. Purely resistive lighting loads, such as incandescent lights, will have a ballast-energy use factor of one. You will

LIGHTING

Location and Equipment name	W	Fixtures	Lamps per Fixture	Est. hours per year lights used	Est. hours per year area used	A/C (Y/N)	Estimated ballast factor	kW[a]	KWh[b]	Estimated Annual Cost ($/yr)[c]
Lighting										
Office										
Lighting line item										
Conference room, 8' fluorescent	60	2	2	2,025	500	Y	1.15	0.28	559	54
Lobby and hallway, 8' fluorescent	60	4	2	2,025	2,025	Y	1.15	0.55	1,118	108
Office lighting sub-total:								***0.8***	***1,677***	***162***
Saw Mill										
Saw mill, 8' fluorescent	95	7	2	2,025	2,025	N	1.15	1.53	3,097	251
Saw shack, incandescent	100	1	1	2,025	2,025	Y	1.00	0.10	203	16
Head saw, halogen spotlight	95	1	2	2,025	2,025	N	1.00	0.19	385	31
Saw mill, HPS	150	8	1	2,025	2,025	N	1.15	1.38	2,795	226
Saw mill lighting sub-total:								3.2	6,479	524
LIGHTING TOTALS:								**4**	**8,156**	**687**

[a]kW = kW/lamp x Number of Fixtures x Number of Lamps per fixture x (Ballast factor + 1)
[b]KWh = total kW x Estimated hours per year
[c]Estimated Annual Cost = kW x $/kW/month x 12 months/yr + kWh x $/kWh

MOTORS

Location and Equipment Name	HP (rated)	units	hours per year	lf	uf	df	Est. Efficiency	Number of Belts	kW[a]	kWh [b]	Estimated Annual Cost ($/yr)[c]
Saw Mill											
Motor line item											
Chipping edger	75	2	2,025	0.4	0.6	0.9	0.94	8	42.81	57,793	5,701
Head saw	100	1	2,025	0.4	0.6	0.9	0.95	4	28.42	38,366	3,785
Log trimmer	15	1	2,025	0.4	0.6	0.9	0.91	-	4.43	5,976	590
Conveyor	5	2	2,025	0.4	0.8	1.0	0.88	-	3.41	5,525	496
Saw mill motors sub-total:									***79***	***107,660***	***10,571***
Miscellaneous											
Drying room fans	2	8	8,760	0.9	0.6	1	0.84	1	12.79	67,217	4,465
Drying room fans	7.5	6	8,760	0.8	0.6	1	0.90	1	30.01	157,715	10,478
Miscellaneous motors sub-total:									***43***	***224,932***	***14,943***
MOTORS TOTALS:									**122**	**332,592**	**25,514**

[a]kW = (Hp/unit x number of units x lf x 0.746 x df)/ Efficiency
[b]KWh = kW x hours per year x uf / df
[c]Estimated Annual Cost = kW x $/kW/month x 12 months/yr + kWh x $/kWh

Fig. 1. Electrical equipment and energy balance (sample audit—sawmill and office building).

AIR-CONDITIONING

Location and Equipment name	Tons	units	SEER	df	hours per year	kW[a]	kWh[b]	Estimated Annual Cost ($/yr)[c]
A/C line item								
Scragger booth, window unit	1	3	8	1.0	2,000	4.50	9,000	733
A/C sub-total:	3.0		8.0					
Average COP:			2.3					
A/C TOTALS:						**4.5**	**9000**	**9,005**

[a]kW = tons/unit x number of units x 12 x df / SEER
[b]KWh = kW x CDD / df
[c]Estimated Annual Cost = kW x $/kW/month x 12 months/yr + kWh x $/kWh

AIR COMPRESSORS

Location and Equipment Name	HP (rated)	units	hours per year	lf	uf	df	Estimated Efficiency	Number of Belts	kW[a]	kWh[b]	Estimated Annual Cost ($/yr)[c]
Air Compressors line item											
Maintenance shop	15	1	2,025	0.4	0.8	1.0	0.91	-	4.92	7,968	869
Saw mill	20	2	2,025	0.4	0.8	1.0	0.91	3	13.12	21,249	1,908
AIR COMPRESSOR TOTALS									18	29,217	2,777

[a]kW = (Hp x number of units x lf x 0.746 x df)/ Efficiency
[b]kWh = kW x hours per year x uf / df
[c]Estimated Annual Cost = kW x $/kW/month x 12 months/yr + kWh x $/kWh

The Energy Balance

CALCULATED FIGURES				**148**	**378,964**	**37,983**
ESTIMATED MISCELLANEOUS	**0.1**			**14.8**	**37,896**	**3,798**
TOTALS				**163**	**416,861**	**41,781**
MIN ACTUAL kW				**145**		
MAX ACTUAL kW				**165**		
ACTUAL kWh | Cost					**416,900**	**41,700**
ERROR IN kWh | Cost				**-1.1%**	**-0.01%**	**0.19%**

Fig. 1 (*Continued*)

have to take into account the ballast factor for fluorescent lights and HID lamps. If you cannot obtain information that is more accurate, you can estimate the ballast-energy use factor to be 1.15–1.20, meaning that the ballast consumes 15%–20% energy in addition to the rated power of the lamp. See the lighting section of Fig. 1: Energy Balance—Electrical Equipment.

- *Specific process equipment*. Sometimes the facility process will have specialized equipment that should be examined carefully to make sure all energy data is gathered. For example, a plastic-products facility may have injection molding machines. These machines will have motors, which should be included in the motor section with appropriate load factors. They will generally have heaters that heat the plastic before it enters the injection process, and they may have some type of chilling (typically chilled water) to cool the final product. All these energy-using components of the injection molding process must be recorded in the data collection process. They can be entered in a separate injection molding process equipment section of the energy balance if desired.
- *Gas and oil powered equipment*. Nameplate ratings in British thermal unit per hour, efficiencies, and load factors must be obtained. You should also verify the maintenance schedule for possible tuning opportunities.
- *Miscellaneous equipment*. You will not need to collect detailed data on small energy consumers such as office equipment or other plug loads at a manufacturing facility. We typically consider the miscellaneous uses to account for 10% of the actual kW and kWh use in a facility. That number goes into our energy balance as part of the total facility energy use. However, in a commercial facility or office building, the computers, fax, and copy machines may account for a substantial energy use, and do need to be recorded.

Data Compilation

After collecting the data for the energy-using equipment, you need to record and organize the data and simulate the operations of this equipment to develop an estimated energy use to compare to the historical energy use. At the UF IAC, we used an excel spreadsheet to organize the data in preparation for data analysis. We started with a spreadsheet template that already contained the basic layout and equations. To prevent inadvertent altering of the equations, we locked the fields containing equations.

We typically organized the data in one of four ways:

- By type of equipment;
- By building or location;
- By process;
- By utility meter or other energy use measuring device.

These different organizational methods are not mutually exclusive. Often, we used a combination of these methods to help us focus on the energy-using equipment. For example, if we organized by building, we would have a sub-organization by type of equipment. We might also organize by utility meters where the facility had more than 1 m, then by equipment on that meter. No matter how you decide to organize your data, you should sub-total the energy use, demand, and cost of each area. You will find these sub-totals useful in latter analysis. Fig. 1: Energy Balance—Electrical Equipment contains the formulas used in each section of the energy balance and shows a sample layout of the sections of an energy balance.

By Equipment

Our fundamental organization is always by type of equipment. For example, we typically start by listing lighting, process motors, heating ventilation and air-conditioning equipment, air compressors, and specialized process equipment. We list any non-electric power equipment, such as natural gas or oil-fired boilers, separately. (Natural gas-fired equipment will also be separated by meter.) Then, we assign the power ratings, hours of use, and initial load, utilization, and diversity factors to each piece or group of equipment. Fig. 1 illustrates this basic layout for the electric portion of an energy balance.

By Building or Location

You may also find it useful to further separate the equipment by building or location. One obvious example would be to separate lighting by office, parking lot, and manufacturing area because the lighting is used differently in each area. The office may only be operated from 8 to 5, while the manufacturing area operates 24 h/day. The parking lot lighting should only operate at night. We found that this type of separation can help identify energy-saving opportunities. For example, if the parking lot lights were operating more than the nighttime hours, then it might be profitable to install photo-sensors. The lighting section of Fig. 1 illustrates the office and a saw mill.

By Process

For a facility that has multiple processes or multiple lines performing similar processes, you may find it useful to further separate equipment by process. Different processes may have different operation schedules, and their separation makes it easier to understand how energy is used in each. For example, a lumberyard may cut lumber 8 h a day, but it may take 60 days to dry the lumber, and the drying processes may operate 24 h/day. Once the data is organized in this manner, opportunities for improvement may become obvious. Does the lumberyard need a better method for drying wood to keep up with the cutting operations? In another example, we visited a plant that had multiple lines performing similar processes. After we had organized the equipment by lines and analyzed the diversity factor, we found that their peak load could be lowered by managing the maintenance and line change-over schedule.

It is very helpful to consider the whole process when you are checking to see if your equipment list is complete. For example, if the facility spray-painted items that went on a conveyor line and through a drying oven, you would check to see if you had recorded an air compressor (for spray painting), conveyor motors, and heating elements.

By Meter

The more meters a facility has, the more accurately you can simulate the energy use at that facility. If a facility has multiple meters, develop a separate energy balance for each meter. Then, at the end, aggregate the values to get the total facility usage. For example, the office areas may have a different electric meter from the manufacturing plant. You may find it useful to first separate by meter, since each meter will have a different bill. Of course, you must separate gas-metered equipment from electric-metered equipment in any case. In addition, you will find it necessary to separate equipment related to each different type of energy sources: oil, coal, wood, propane, etc. Fig. 1 shows only the electric use section of the energy balance. We use the same energy balance template for other meters or energy types by using a similar section with the appropriate equations and units for that energy source and that equipment.

VALIDATING ENERGY BALANCE WITH HISTORICAL USAGE PATTERNS

Once you have collected and organized the data and entered it into the energy balance spreadsheet, start by totaling the demand and energy use from each equipment sector to get a total for the facility. Then you must compare it to the historical totals. If you have separated the equipment by meters, you should perform this balance for each meter and then aggregate each part at the end.

If your data collection and utility measurements are accurate, you should have a reasonably good estimate of the energy use by equipment sector. If the energy balance values are fairly close to the historical usage, you can fine-tune the energy balance values using the steps described below (see Fine-tuning the Energy Balance). However, if the values are more than a few percentage points apart, you should check the following areas before you start trying to adjust the energy balance.

- Was equipment missed during the site visit?
- Did you fully understand how the equipment is used?
- Did you fully understand equipment diversity?
- Did you fully understand the motor load factors?
- Did you fully understand the utility bills? Did you record them correctly?
- Was the energy demand and use metered correctly by the utility?

Comparing Energy Balance to Historical Energy Bills

The past use of energy and power is a good indicator of how a plant may be currently using their energy-consuming equipment. However, you must be aware of any changes in operation between the past and present. For example, you must consider if the plant has added any equipment or changed production schedules when you use the historical energy bills. Again, the goal is to specifically account for 90% of the historical energy use, and a demand between the actual minimum and maximum demand values achieved during the last business cycle. If a large piece of equipment has been added or removed during the period for which you are analyzing the bills, you must adjust the hours of use for that equipment for the appropriate period.

Sources of Problems with Reconciling the Energy Balance to Historical Energy Bills

You will find that at times you have trouble reconciling your predicted use with the historical use. In these cases, we have found a number of reasons for this inability to satisfactorily complete the energy balance. An advantage of performing the energy balance in the manner we describe is that this process may uncover a problem either of your understanding how a plant uses energy or a problem that must be addressed by the plant or the utility company.

Missing Equipment from the Plant Audit

Perhaps during the plant visit, you missed collecting data on some large energy-consuming equipment. The best way to approach this issue is to verify with plant personnel that you covered all primary and secondary processes at the plant. Sometimes it is easy to miss some of the less significant processes at a plant. The plant may produce a by-product from its processes that are almost an afterthought for plant personnel. For example, they may not have realized that processing scrap requires energy use at the plant. Equipment required for processing scrap may include shredders, compactors, cyclone separators, etc. Maybe you made the plant visit in the winter and the company uses portable coolers or a large number of personal fans in the summer for employee comfort.

Faulty Understanding of Equipment Usage

Possibly, during the plant visit, you did not properly find out how or when equipment is used because plant personnel may not understand its use. Equipment run primarily during off peak business hours may be a prime suspect. For example, plant personnel may give you faulty information about how personnel on shifts, other than his shift, actually use the energy-consuming equipment. Maybe, plant personnel who are on site during the day do not know when outdoor lights are used. To approach this issue, you could interview plant personnel or make follow-up visits during various times to get a better understanding how the equipment operates. For example, we came across one plant that we estimated a higher energy use than the historical numbers showed was possible. We over-estimated the energy use because initially the auditor misunderstood use of the plant-electric heaters. All the heaters combined to account for a large power use. Initially, the auditor assumed a use factor that was too high. Actually, the heaters were only used on the coldest nights of winter, and then they would cycle with only a portion being energized at any given time. The exercise of creating the energy balance helped the auditor understand how the plant used their heaters.

Faulty Understanding of Equipment Diversity

A 1-day visit to a plant is a small sample. The day of the visit may not have been a typical day. Observing that two identical pieces of equipment do not operate at the same time may be an aberration. Unless there is control equipment that prevents the operation at the same time, you cannot be sure. To approach this issue, you should interview plant personnel about the interactions between the equipment, so you can estimate the diversity of equipment operations. Again, you could also make periodic visits to observe how the equipment actually operates even if these visits are on the same day during your audit visit.

Faulty Understanding of Load Factors

You need a good understanding of the basic ideas of typical equipment design (or over-design). To approach this issue, you could take some sample measurements of equipment for large energy-consuming equipment to get an idea of the load factors.

Faulty Understanding of Utility Bills

Sometimes, some of the equipment will be separately metered, and you did not receive the bills for these meters. If you check the bills with the utility representative, you may understand the utility billing structures better than the plant personnel may. An alternative would be to interview the plant employee who pays the bills to make sure you have them all.

Faulty Measurement of Energy Use by the Utility Meter

Sometimes with the merger and divestiture of companies, a plant may be sold to two different parties. We found one such case where the plant meters were not properly divided. Therefore, one plant was unknowingly giving electricity away to another plant with a different owner. Sometimes meters are not working properly or meter readers have read them wrong. These problems need to be brought to the attention of company management and the utility company, so the problem can be solved.

FINE-TUNING THE ENERGY BALANCE

If the estimated demand and energy use do not match the historical demand and energy use within 2%, including miscellaneous, you must begin revising your assumptions to reconcile the collected data on energy-using equipment with the historical data. Often we were able to match historical energy use to within fractions of a percent. Once you have complete confidence in the data you have collected, then the only parameters that you can adjust or "tweak" are the load, utilization, and diversity factors. Again, the goal is to account for 90% of the historical demand and energy consumption for every type of energy. The other 10% you can consider caused by miscellaneous loads. By adjusting these three factors, you can develop a fairly accurate balance of the energy and demand use at a manufacturing plant.

How to Adjust the Energy Balance

There are eight possible mismatches between the total calculated facility demand (kWc) and energy use (kWhC) and the historical demand (kWh) and energy use (kWhH). You must adjust the load, utilization or diversity factors depending on which situation you have. As you adjust one of the parameters, you may find that you then have to adjust another to fine-tune your results. Table 1 shows the adjustment to make for each the situation.

Adjusting the Load Factor

As previously mentioned, we have found that most process motors have a load factor between 0.35 and 0.6. We generally start at 0.4. Some motors such as the air compressors and fans will have higher load factors typically from 0.8 to 0.9. If your total kW and kWh are a little too low

Table 1 Adjusting the load, utilization, and diversity factors

Situation	Demand	Energy	Adjustments
1	kWc < kWH	kWhc < kWhH	Increase the load factor on some of the motors
2	kWc > kWH	kWhc > kWhH	Decrease the load factor on some of the motors
3	kWc = kWH	kWhc < kWhH	Increase the utilization factor on some of the equipment
4	kWc = kWH	kWhc > kWhH	Decrease the utilization factor on some of the equipment
5	kWc < kWH	kWhc = kWhH	Increase the diversity factor on some of the equipment
6	kWc > kWH	kWhc = kWhH	Decrease the diversity factor on some of the equipment
7	kWc > kWH	kWhc < kWhH	Decrease the diversity factor and increase the utilization factor on some of the equipment
8	kWc < kWH	kWhc > kWhH	Increase the diversity factor and decrease the utilization factor on some of the equipment

(or too high), then increase (or decrease) the motor load factor slightly for the lightly loaded motors. However, if the total kW is significantly higher or lower than the historical demand, then you will have to look for some other reason (see Validating Energy Balance with Historical Use above). To change the energy and power use for a motor on the energy balance, you can change the load factor.

Adjusting the Utilization Factor

As discussed previously, we chose the annual hours of use for a piece of equipment as the operating hours for the plant or department. Then, we used the utilization factor to account for cycling of equipment, maintenance downtime etc. If your kWh is too low or too high, you can adjust the utilization factor to help reconcile the difference. You might want to adjust it for specific pieces of equipment rather than an across-the-board adjustment on all equipment. Changing the utilization factor changes only the estimated annual energy use for a piece of equipment.

Adjusting the Diversity Factor

We used the diversity factor to calculate the average demand when two (or more) similar pieces of equipment are highly unlikely to run at the same time. For example, if a plant had two pumps, but when one was running, the other was in standby mode, then the diversity factor for each pump would be 0.5. Another example would be a plant that has a large number of fans, but the probability of a fan running at the time of the peak was 75%. Thus, the diversity factor of each of the fans would be 0.75. Changing the diversity factor changes only the estimated annual demand for a line item on the energy balance for a piece of equipment.

DATA ANALYSIS

Once you are comfortable that your data represents the actual energy use (i.e., your calculated demand and energy use accounts for 90%, excluding miscellaneous, of the historical use), you can use the energy balance spreadsheet to produce several useful graphs that show the relative energy use of each set of equipment (lighting, motors, etc.), the relative energy use of groups of equipment by each separately metered area, and so forth. We used a pie chart to show the relative usage and cost of each type of electrical equipment. This pie chart gave us a good indication of the areas with the greatest opportunity for energy savings. It also allowed us to show the plant personnel where their energy was being used. In addition, viewing the pie charts helped UF IAC auditors with less experience gain a better feel for the areas to concentrate upon during subsequent audits for similar plants. See Fig. 2 for a sample pie chart based on our sample energy balance.

EXAMPLES OF BENEFITS RECEIVED USING AN ENERGY BALANCE

At the UF IAC, we have had many instances when the energy balance method described here has aided us with understanding the energy use of a facility and with recommendation ideas. Here are three examples of these instances.

Interactions between Recommendations

During the analysis of recommendations for a client, we made a series of interrelated recommendations including high-efficiency motors, demand shifting, and synchronous

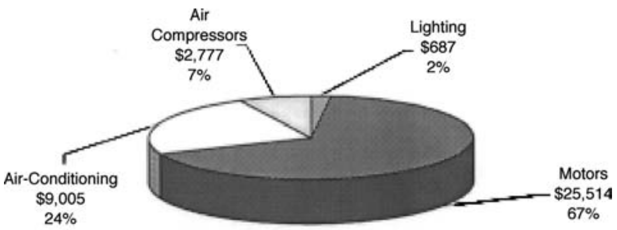


Fig. 2 Estimated cost of operation by type of electric device.

v-belts. Separately, these recommendations showed a total energy savings of about 50%, which seemed highly unlikely. After reviewing our energy balance, we realized that these recommendations were not mutually exclusive. A before-and-after energy balance showed that the actual savings from the combined measures was much less than 50%.

In another case, we prepared a series of lighting recommendations: T-8 lighting, occupancy sensors, photo-sensors, compact fluorescent lighting, and lower wattage halogen lamps. However, the combined savings would have saved more than the energy balance predicted that the lighting system used in total. Again, using the before-and-after energy balance pointed out the result of double-counting the savings.

Our most memorable example of the need for an energy balance was in our early audits when we used an initial motor load factor of 80% for a facility and calculated a savings in demand and energy that was higher than the total historical demand and energy use for the facility. At this point, we realized that we needed a tool for better analysis, and this gave rise to the development of our energy balance method of energy and demand analysis.

CONCLUSION

Before you can save money through the efficient use of energy, you must understand how that energy is used. One strategy would be to monitor equipment with an energy management system or some other monitoring method. Monitoring takes a period of observation and is costly. Even if the expense of this costly monitoring is undertaken, it is unclear if you will benefit from it without significant additional analysis. With the method we describe in this article, you can gain many of the benefits of the more costly methods with less cost. Maybe, after performing this method, you will find that installing an energy management system will yield appropriate benefits, or maybe you will find that the more costly methods will not yield an acceptable return on investment. Therefore, depending on the situation, the energy balance method may provide a relatively low-cost and reasonably accurate alternative to understanding the energy use at a facility.

REFERENCES

1. Capehart, B.L.; Capehart, L.C. *Improving Industrial Energy Audit Analyses*. Proceedings of the ACEEE Energy Efficiency in Industry Conference, Grand Island NY, Summer 1995.
2. MotorMaster and the Motor Challenge, http://mm3.energy.wsu.edu/mmplus.
3. Capehart, B. *Equipment Load Factors, Use Factors and Diversity Factors as well as a General Discussion of Energy Audit Procedures*, http://www.ise.ufl.edu/capehart/papers/diversity.doc, 2000.

Facility Energy Use: Benchmarking

John Van Gorp
Power Monitoring and Control, Schneider Electric, Saanichton, British Columbia, Canada

Abstract
Energy management practices have evolved significantly over the past decade, in many cases borrowing from advances in business management methods and information technology. One key aspect of this comprehensive approach to energy management is a focus on setting goals and measuring performance against these goals, a process that is often referred to as benchmarking.

An energy benchmark is a reference used to compare the performance of one energy system against another. This article describes several key concepts in the construction and use of energy benchmarks, including selection of data sources, data management concepts, constructing baseline benchmarks and analysis tips. Examples are used throughout the article to illustrate how benchmarks can be used to manage facility energy use.

INTRODUCTION

"What gets measured, gets managed"—Peter Drucker, business management theorist

Energy management practices have evolved significantly over the past decade, in many cases borrowing from advances in business management methods and information technology. The traditional approach to energy management focuses mainly on replacing existing equipment with new equipment that takes advantage of the latest energy efficiency technology. Modern practices extend beyond this focus on energy technology and take a comprehensive approach to the process of managing energy.

The ANSI/MSE 2000 standard[1] and the Energy Star Energy Management Guidelines[2] are examples of this more comprehensive approach to energy management. Both of these guidelines provide a framework for setting energy consumption goals, tracking performance against those goals, and communicating results. Both also emphasize the critical role that measurement and verification plays in the success of most energy management programs. One U.S. Department of Energy paper studied energy efficiency projects at more than 900 buildings and found that projects that implemented best practices in measurement and verification realized higher savings (both initially and over time) than comparable projects, yielding an additional return on investment of nearly 10%.[3]

One key aspect of this comprehensive approach to energy management is a focus on setting goals and measuring performance against these goals. The process of setting energy goals requires an understanding of current energy use and comparison of that use against one or more reference standards. This process is normally referred to as benchmarking.

Energy Benchmarks

An energy benchmark is a reference used to compare the performance of one energy system against another. Energy benchmarks are often used to compare the energy consumption of different types of equipment (such as motors and lights), buildings, and industrial processes. The benchmark used can either be internal or external, as described below:

- An internal energy benchmark is a reference built by an organization to compare similar energy systems within the organization. This comparison is normally used to highlight the most efficient energy systems and isolate the positive behaviors that drive this efficiency.
- An external energy benchmark is a reference built by industry associations and government organizations to compare similar energy systems. This reference is normally obtained through analysis of many similar energy systems operating in comparable environments. The best practice behaviors that drive the most efficient energy systems are often documented to help the organizations using the benchmark increase the performance of their own energy systems.

To create a true and accurate comparison, most energy benchmarks are normalized by some key driver of energy consumption, and important aspects of the environment that might affect the benchmark are stated. This results in a measurement reference that is common to both things being compared. Benchmarks for building energy use are commonly normalized by floor space, and benchmarks for industrial energy use are commonly normalized by some

Keywords: Energy benchmark; Energy analysis; Energy monitoring; Energy management; Measurement and verification.

unit of production. Environmental aspects that might affect the application of a benchmark include the major use of a building, occupancy, and climate.

To help illustrate this concept, some normalized energy benchmarks commonly used are:

- Building energy consumption: kWh per square foot (or per square meter).
- Manufacturing energy consumption: kWh per ton of production.

Government organizations, such as Energy Star, Natural Resources Canada, and Rebuild America, offer energy benchmarks for a variety of applications. Industry associations, such as the Canadian Pulp and Paper Association (CPPA), offer energy benchmarks relevant to members of their industries. Example energy benchmark guides include the following:

- *Benchmarking Your Facilities for Greater Success* from Rebuild America[4]
- *Benchmarking and Best Practices Guide for College Facility Managers* from Natural Resources Canada[5]

Some organizations may prefer to express an energy benchmark in terms of cost. A manufacturer, for example, might use both kWh-per-ton and energy-cost-per-ton benchmarks to track performance. Expressing a benchmark in terms of energy cost can be useful for cost accounting purposes, but the rate structure for many energy users can make such a calculation difficult. Such a benchmark should not be used for managing the efficiency of an energy system. Changes unrelated to energy consumption (such as changes in the energy rate charged by the local utility) can affect this metric.

Preparing for Benchmarking

Benchmarking preparations are described in the following steps:

- Benchmarking scope
- Selecting data sources
- Data management
- Constructing baseline benchmarks

Benchmarking Scope

The scope of a benchmarking exercise depends primarily on the energy management goals that will be supported. If the goal is simply to assess the current state of energy efficiency within an organization, or to implement a straightforward energy efficiency project, then benchmarking preparations will not be as extensive as they would be for a benchmarking effort designed to support a comprehensive energy management program. Data collection, monitoring, and reporting efforts for one-time benchmarking exercises normally require much less time (and expense) than efforts designed to create an active and ongoing benchmarking program.

Active benchmarking programs, however, can often help drive greater energy savings over a longer period.

Data Sources

Once the scope of the benchmarking effort has been determined, the next task is to find sources of the data required to build up the benchmarks. These data can be organized into two main categories:

- Static data such as equipment ratings and floor area. These are data that tend to change slowly (or not at all) over time.
- Dynamic data such as energy consumption and production volume. These data change frequently and need to be collected at regular intervals.

Most of the expense and effort involved in data management is due to dynamic data collection and processing. Collecting static data, such as floor area, can often be performed just once; dynamic data, such as energy consumption, on the other hand, requires a continuous effort. The number of dynamic data measurements should normally be kept to the minimum required to support energy management goals. Once these measurements have been determined, the time interval used for data collection should be short enough to provide the detail required to understand patterns of energy consumption. More than this level of detail will increase the cost of data management without resulting in a greater understanding of energy consumption.

Sources of dynamic energy consumption and normalizing data include the following:

- *Energy consumption from utility bills*. Energy consumption data can be manually entered from utility bills, but many utility companies will also offer billing data in an electronic format to their larger customers.
- *Energy consumption from "shadow" metering*. If detailed energy consumption data is required but not available from the utility company, a facility can install a separate meter at the utility service point to "shadow" the utility meter.
- *Energy consumption from sub-meters*. Benchmarking equipment or processes within an energy system may require the installation of separate meters to capture the data required.
- *Energy consumption from existing automation systems*. Some building and industrial automation systems can integrate with basic energy meters and collect energy consumption data.

- *Production data from existing automation systems.* Many manufacturing organizations record production volume using some form of information system, ranging from process historians in process control systems to shipment data in material resource planning (MRP) systems.
- *Temperature and humidity data.* Sources for this data range from a variety of publications and online weather services and to on-site weather stations.

Electricity and natural gas are the most common forms of energy monitored in benchmarking programs, but the information presented here can be applied to other forms of energy transfer as well (such as steam or chilled water). When aggregating or comparing different forms of energy, it is important to convert measurements to common units (such as BTUs) first. Energy management references, such as the *Energy Management Handbook*,[6] list conversion factors and instructions on their use.

Data Management

The next step after selecting which data needs to be collected and the source for this data is how this data will be managed, processed, and communicated. The choice of which data management approach to follow depends primarily on the energy management goal being supported. If the benchmarking exercise is intended to be a one-time comparison against internal or external benchmarks, then manual data entry into a spreadsheet may be sufficient. If benchmarking will be supporting an active energy management plan, and a large number of energy measurements are involved, then a more comprehensive energy information system will likely be required.

Here is a summary of both the spreadsheet and energy information system approaches:

- *Spreadsheet with manual data entry.* Energy consumption information obtained from utility bills or read directly from meters is manually entered into a spreadsheet. Simple calculations are used to divide energy values by the key normalizing value (such as floor area) to generate the benchmark measurement. This step is repeated for each energy system being monitored.
- *Enterprise energy management system.* Digital energy meters and manual data entry transmit measurements to a central server that archives and processes the data. Benchmarks are automatically calculated for all energy systems being monitored. Current and historical results are available as tables and graphs easily accessible by multiple users.

In addition to automating data collection and benchmark calculations, an energy information system can also collect and organize more detailed interval data for analysis of energy consumption patterns. If benchmarks indicate a sudden increase in energy consumption, this interval data can be crucial to understanding the drivers behind this sudden increase.

Fig. 1 shows the typical components that form a modern energy information system. Microprocessor-based devices measure energy use at key points within one or more facilities and communicate this data back to a server through a communications network. The software archives this data, processes it as required, and presents it to users in a variety of ways. Information can typically be accessed using a standard web browser or by receiving alert messages through wireless devices.

Constructing Baseline Benchmarks

When constructing an energy benchmark, an organization must select energy consumption data over some length of time. This baseline data set is selected to describe the current performance of an energy system. Data should be collected over a period of time that is long enough to capture nearly all of the operating modes of the energy system (through different seasons of the year or different levels of production, for example). The period selected for the baseline should not include major changes to the energy system (such as major equipment retrofits).

The example below illustrates the steps involved in building energy benchmarks.

Example 1. Energy benchmarks for a portfolio of office buildings.

A property management company with a portfolio of five office buildings suspects that some buildings are consuming much more energy than they should. A review of electric utility bills for all buildings over the past year indicates that consumption varies by a factor of five between the smallest and largest buildings!

An energy manager at the property management company has been given the task of benchmarking the energy efficiency of all five buildings against each other and against a recognized reference benchmark. The goal is to identify which buildings are the most and least energy efficient and to compare all buildings to the recognized reference benchmark. This benchmarking exercise will be the starting point towards understanding current energy use and launching a comprehensive energy management plan across the organization.

The energy manager takes the following steps to create energy benchmarks for the five office buildings:

- *Acquire electrical energy consumption data and normalizing data.* The energy manager obtains copies of the electric bills for each of the five buildings for the previous year. She creates a new spreadsheet and manually enters the consumption data for each building for each month of the year, starting with January. She

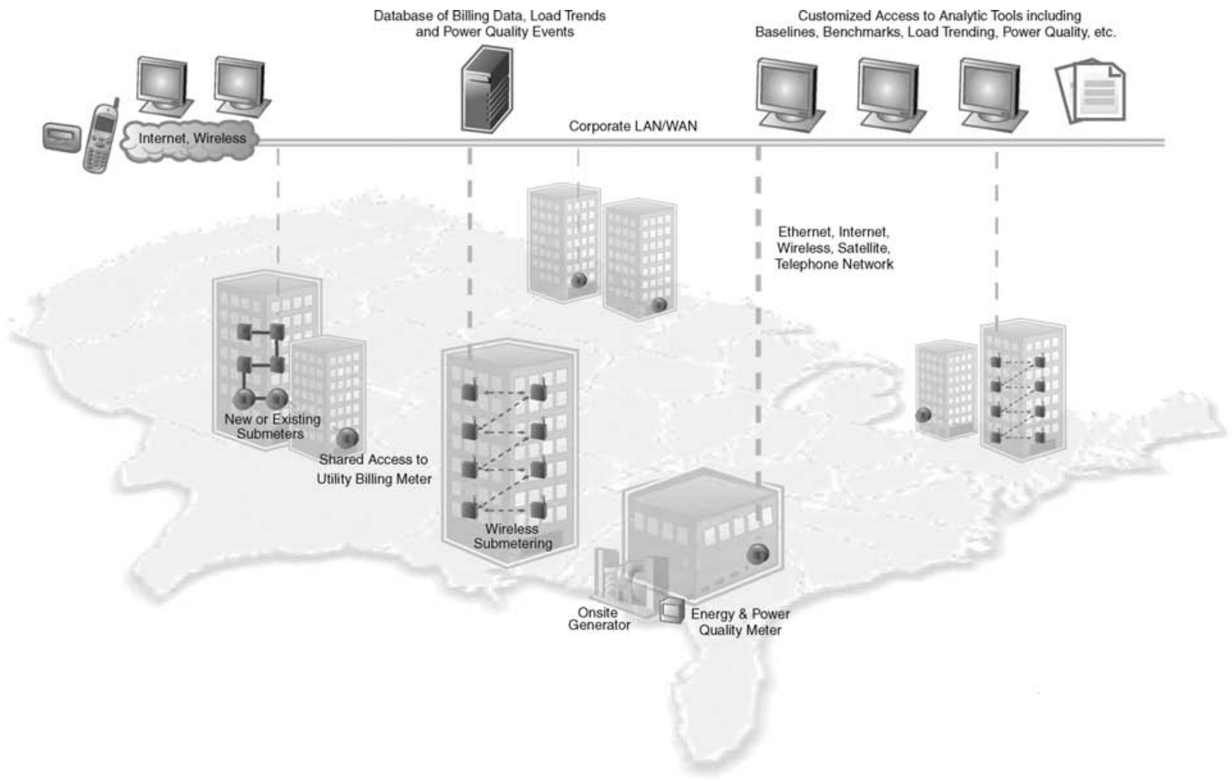


Fig. 1 Components of a typical enterprise energy management system.

also enters the total floor space for each of the five buildings.

- *Select an appropriate reference benchmark*. The energy manager decides to use reference benchmark values from the "Benchmarking Your Facilities for Greater Success" document from Rebuild America.[4] This document provides electrical energy consumption benchmarks for a variety of building types, based on data from a U.S. Department of Energy study in 1999. The energy manager decides to use the median annual benchmark value for office buildings, which is 11.7 kWh/ft^2.
- *Calculate annual total kWh/square foot benchmark values*. The monthly kWh consumption values for each building are totaled to get an annual consumption total. This total is then normalized by dividing by the floor area for the associated building to get annual kWh/ft^2 benchmarks for each of the five buildings.
- *Set baseline benchmark values for each building*. The benchmarks calculated in the step above are recorded as the baseline benchmarks for each building. These values will be used to track changes in energy efficiency over time.

Fig. 2 below compares the annual kWh/ft^2 benchmarks for each of the five buildings against each other, and against the Rebuild America reference benchmark. Buildings 1–3 consume more energy per square foot than the reference benchmark, and buildings 4 and 5 consume less. Initial interest in this energy benchmarking exercise arose when a review showed that the largest building consumed much more energy than the smallest building, but the energy manager is surprised to see that the benchmark comparison tells a different story. The largest building (building 4) is actually one of the most energy efficient buildings in the property management company's portfolio, and the smallest building (building 2) is one of the least efficient!

This example focused solely on monthly electrical energy consumption data from utility bills and a benchmarking exercise to examine where the buildings currently stand. This example also used a simplified approach to benchmarking by excluding factors that can influence energy consumption, such as outdoor temperature and manufacturing production volume. The next section describes how such benchmarks can be used on a continuous basis to support an energy management program.

Benchmarking in Action

Although benchmarking may first be approached as a one-time exercise to determine the energy efficiency of a facility, it delivers the most value when practiced on a

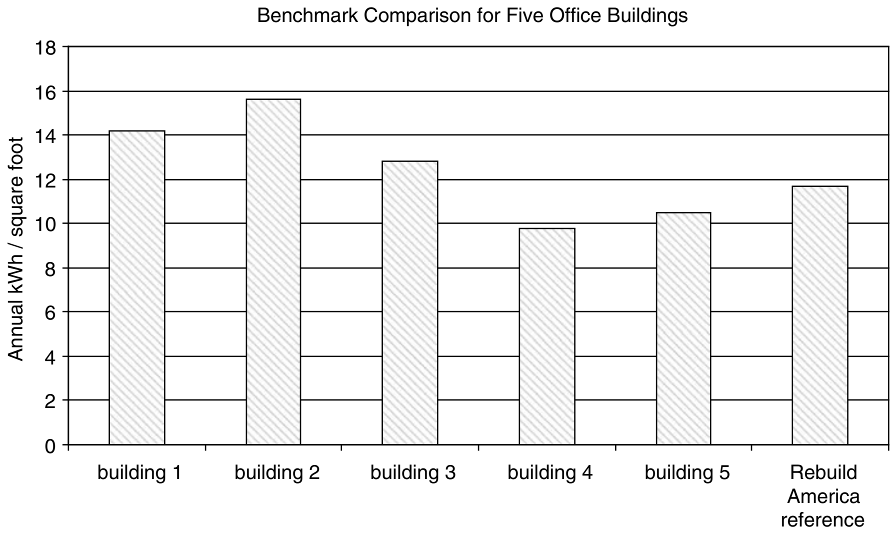


Fig. 2 Benchmark results for five office buildings.

continuous basis as part of a comprehensive energy management program. Energy benchmarks can provide a valuable summary of the performance of a variety of energy systems used throughout an organization. Different information users within an organization can be given the energy data they need in a manner that will help them support key energy management goals. Energy benchmarks can also be supported with additional drill-down details that help experts use the information to understand the drivers behind changes in energy consumption.

The sections below describe how energy benchmark information views can be configured for a variety of users to give them the information they need to support the success of an energy management program. The last section provides several tips designed to help energy managers uncover energy use patterns and get the most out of their benchmarking data.

Determine Benchmark Users

The first step in designing an energy benchmarking information system is to determine who will be using it, and how they will be using it. A variety of people within an organization can benefit from benchmarking information and use this information to support active energy management projects.

Some user roles and their typical energy benchmark information requirements include the following:

- *Executives*: Executives normally require only high-level benchmark summary data that shows energy efficiency performance against set goals. Executives may review benchmark summary reports once per quarter. If actual energy consumption is significantly higher than the target goals in an energy management plan, executives may request a more detailed report that explains the reason for a variance from planned consumption.
- *Facility engineering staff*: Facility personnel often require aggregated summary views backed by detailed time-series data that shows daily or hourly energy consumption and demand data. Engineering staff may review aggregated summary views once a week. If there are significant increases from energy consumption goals, engineers may review summary views and detailed time-series data once a day and take the actions necessary to put energy consumption back on track.
- *Tenants*: Tenants in retail or office space can receive monthly energy benchmark reports along with their utility bill. These reports typically compare the tenant's energy consumption against the least efficient, most efficient, and average energy consumption benchmarks for all of the tenants in the facility.
- *Department managers*: Managers that carry energy cost responsibilities in their budgets may receive benchmark information as part of a monthly report showing the performance of their department against budget. This benchmark report will typically show the energy consumption performance of the department against target goals, or compare performance to that of other departments.

Understanding who the users of a benchmarking information system will be provides the guidance needed to design the information views they will use to set energy management goals, track and correct for variances from plan, and proactively seek new opportunities for continuous energy efficiency improvement.

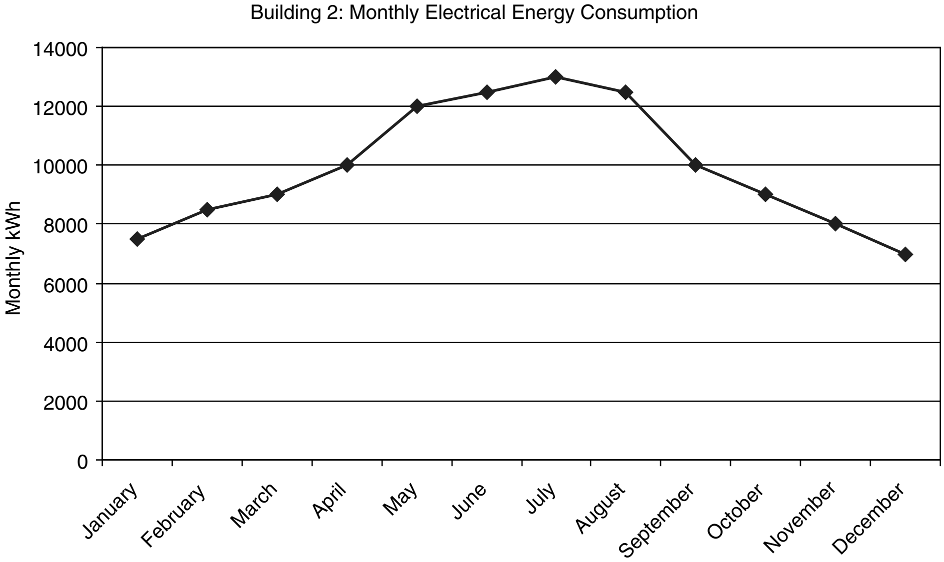


Fig. 3 Monthly kWh consumption values for the least efficient building.

Design Information Views

There are a variety of ways to display energy benchmark information and detailed data. Some example information views include:

- *Energy benchmarks in a table*. A table is often the best way to organize and display detailed energy benchmark values. A table might be used to show specific benchmark values for different buildings, tenants, or departments. If an energy management plan outlines targeted reductions in consumption over some time span, a table could be used to show the yearly energy consumption benchmark values over the time span.
- *Energy benchmarks in a bar chart*. Current and past energy benchmarks can be visually compared against target goals using a bar chart. Such a chart may show month-by-month actual values vs. target values for some energy system. A bar chart can also be used to compare benchmarks for different energy systems.
- *Drill-down data in a time-series chart*. To gain an understanding of what is driving significant increases or decreases in energy consumption, a time-series chart can provide a detailed view of the data behind an energy benchmark. Both actual and target values can be plotted over time to help an energy manager see where any deviations from an energy management plan exist.

Example 2. Drilling down into greater detail

An energy manager for a property management firm has created energy consumption benchmarks for five office buildings, as described in Example 1. Although energy consumption for the largest building is five times larger than the smallest, the bar chart created by the energy manager shows that the largest building (building 4) is actually one of the most efficient of the five buildings, and the smallest building (building 2) is one of the least efficient.

The energy manager decides to drill down into the data behind the energy consumption benchmark value for building 2 to gain a better understanding of its poor energy efficiency. She creates a new worksheet in the spreadsheet she used to calculate the energy benchmarks in Example 1 and enters the monthly kWh consumption values for building 2 from past utility bills. She then creates a time-series graph with kWh consumption on the *y*-axis and months of the year on the *x*-axis.

Fig. 3 shows the resulting time-series graph of the kWh consumption for building 2 over the year. It is easy to see that most energy consumption takes place during the summer months, which the energy manager does not find surprising; many buildings consume more electricity in the summer to run heating, ventilation, and air conditioning (HVAC) equipment to cool the building.

The energy consumed in the summer seems high to the energy manager, so she decides to use the data in her spreadsheet to build monthly kWh/ft^2 benchmarks to compare building 2 to one of the most efficient of the five buildings. She uses data for the months of May through August and normalizes the monthly electrical consumption data for building 2 and building 4 by the area of their floor space. Fig. 4 shows the resulting bar chart comparing energy benchmarks for both buildings for the months of May through August.

Benchmarking Analysis Tips

Once energy benchmarks have been established and standard information views created, most ongoing analysis

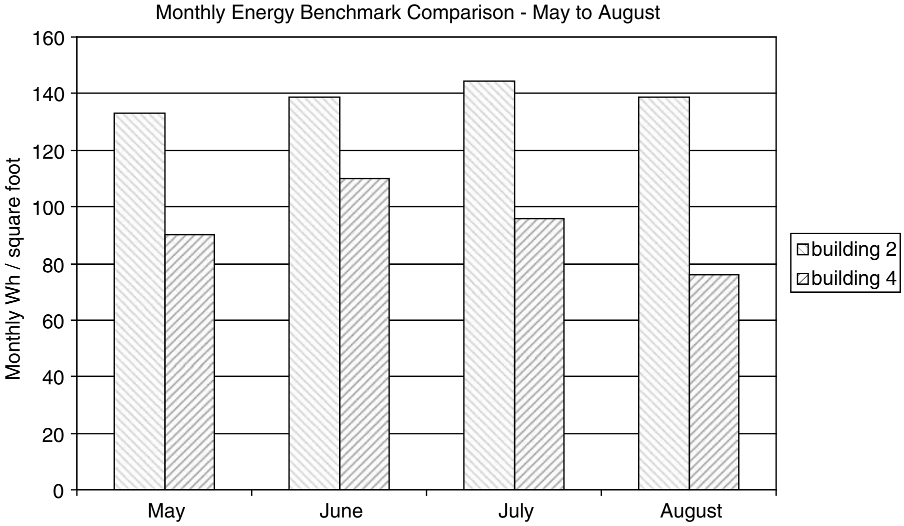


Fig. 4 Monthly energy consumption benchmarks for two buildings over summer.

effort will focus on detecting and understanding changes in energy consumption. The goal of most energy management programs is to first understand and then control energy consumption. The creation of energy benchmarks and baseline consumption profiles are key first steps towards understanding energy consumption. Control of energy consumption is driven by ongoing energy data analysis and taking the steps necessary to gain an understanding of how internal and external drivers are influencing consumption patterns.

One way to uncover hidden patterns of energy consumption is to organize and aggregate consumption data by some key driver of that consumption. Grouping energy consumption by time of use often proves useful in uncovering patterns of energy consumption, especially since time of use can be an excellent proxy for key drivers such as occupancy, outdoor temperature, and production. A few suggestions for time of use aggregations of energy consumption include the following:

- *Seasonal*: Aggregating energy consumption data by season is a simple way to estimate the impact of environmental conditions. Changes in outdoor temperature and daylight hours from season to season can influence the energy consumption of a variety of energy systems (such as HVAC and lighting systems). Example 3 below includes aggregation of energy consumption by season.
- *Weekday/weekend*: Aggregating energy consumption data by weekday and weekend is a simple way to estimate the impact of occupancy. Building occupancy will definitely impact energy consumption, and many buildings (such as offices and schools) have distinctly different hours of operation and levels of use on weekdays and weekends. Example 3 below includes aggregation of energy consumption by weekday/weekend.
- *Production shift*: Industrial energy consumption is primarily driven by production volume, which makes a schedule of production shifts a reasonable proxy for production activity. Manufacturing activity may not change much across seasons or differ by day of the week, but aggregating energy consumption by shifts may uncover interesting differences in efficiency from one shift to the next.

Detecting and determining the drivers behind anomalies in energy consumption patterns is also important to meeting the goals set in an energy management plan. Such anomalies need to be analyzed in order to determine their cause and steps need to be taken to minimize their impact in the future. Two types of anomalies to watch for include sudden spikes and sustained changes in energy consumption, as described below:

- *Spikes*: A sudden spike increase in energy consumption may simply indicate a temporary and infrequent change in equipment operation, such as a seasonal test of a building HVAC system. Such spikes should be investigated to ensure that they are not indicating something more serious, such as the pending failure of energy system equipment.
- *Sustained changes*: A sustained increase or decrease in energy consumption should be expected when there are major changes to an energy system, such as the expansion of a building or replacement of old equipment with more energy efficient equipment.

Unexpected changes should be investigated to determine the cause.

Finally, a greater understanding of energy system behavior can be achieved by creating information views that compare different measurements to one another. Some examples of comparison information views include the following:

- *Compare energy systems*: A common measurement (such as energy consumption) or benchmark for different energy systems is placed in a chart or table for comparison. Fig. 2 is an example of such a comparison. For accurate comparisons, ensure that the units of measure (such as kWh or kWh/ft^2) are the same and that the time range for the data sets is the same.
- *Energy consumption vs. potential driver*: This information view combines a measurement (such as energy consumption) against a potential driver of that measurement (such as external temperature). Such views can provide a quick (though informal) test of a link or correlation between one variable and another; a more thorough test would involve statistical techniques such as regression analysis.

The example below shows how an energy manager might further investigate the root cause of high energy use by organizing energy information and comparing it to a key driver, in this case time of week.

Example 3. Organizing benchmark data to uncover patterns.

To understand why the energy consumption for building 2 seems so high during the summer months, the energy manager contacts the local utility to see if daily energy consumption data for several buildings is available. Her utility account manager confirms that daily interval data is available, and at her request, sends a CD with daily energy consumption data for both buildings 2 and 4 for May–August.

The energy manager creates a new worksheet in the spreadsheet file she has been using so far and imports the daily energy consumption data from the CD. She then creates a new column and uses the "day of week" function to mark days as either weekday or weekend days. Finally, she aggregates weekday and weekend energy consumption for May–August for buildings 2 and 4 to see how they compare. Fig. 5 shows average daily energy consumption for May–August for both buildings, grouped by weekdays and weekends.

The chart in Fig. 5 makes it clear to the energy manager that much of the difference in energy consumption between buildings 2 and 4 is taking place on the weekends. To discover why, she contacts the building managers for both buildings to discuss building occupancy and operation during weekdays and weekends in the summer. Both managers confirm that the majority of their tenants operate between 9 a.m. and 5 p.m. during weekdays and are closed on weekends. The building managers also confirm that building automation systems are in place in their buildings, and that they are configured to operate in a "setback" mode during weekends to conserve energy. The manager for building 2 does admit, however, that the building automation system used in his building is quite old, and no one on his maintenance staff really knows how to operate it. The energy manager describes the results of her analysis so far, and suggests that the automation system may not be operating as expected. She recommends contacting the building automation

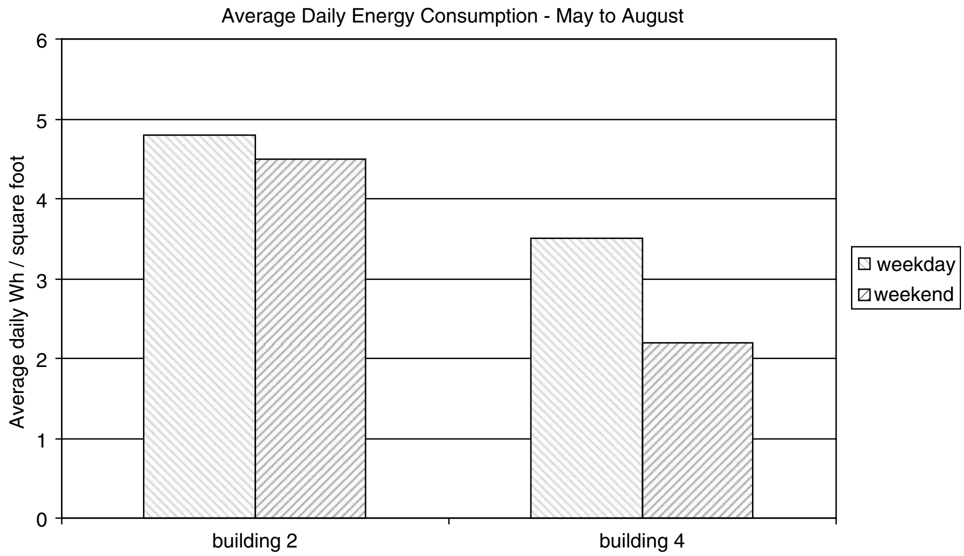


Fig. 5 Benchmark values (weekdays/weekends) for buildings 2 and 4 over summer months.

vendor's services group to confirm the setback configuration and schedule training for the maintenance staff.

CONCLUSION

Energy management practices have evolved significantly over the past decade, in many cases borrowing from advances in business management methods and information technology. In the past, there was a greater focus on equipment technology and less consideration of the management framework required to set and track performance against goals. Modern energy management practices increase and sustain energy savings by adopting the performance management approach integral to modern business and quality management programs.

Information systems that support activities like facility benchmarking are becoming a key part of modern energy management practice. In the past, such information systems were often prohibitively expensive, but advances in measurement and computer technologies are making such information systems cost effective for an increasing number of applications. A comprehensive energy management approach that sets and tracks efficiency goals and incorporates facility benchmarking will help drive increased (and more consistent) energy savings over time.

REFERENCES

1. MSE 2000:2005. *A Management System for Energy*, http://webstore.ansi.org/ansidocstore/product.asp?sku=ANSI%2FMSE+2000%3A2005.
2. *Guidelines for Energy Management, Energy Star Program*, http://www.energystar.gov/index.cfm?c=guidelines.download_guidelines (accessed July 2005).
3. Kats, G.H.; Rosenfeld, A.H.; McIntosh, T.A.; McGaraghan, S.A. Energy efficiency as a commodity: The emergence of an efficiency secondary market for savings in commercial buildings. In *Proceedings of the 1996 ACEE Summer Study*, Vol. 5, 1996, 111–122.
4. Benchmarking Your Facilities for Greater Success, Rebuild America, http://www.rebuild.org/attachments/SolutionCenter/RBABenchmarkingSuccess.pdf (accessed July 2005).
5. *Benchmarking and Best Practices Guide for College Facility Managers, Natural Resources Canada*, http://oee.nrcan.gc.ca/publications/infosource/pub/ici/eii/m92-189-2000e.cfm?attr=20 (accessed July 2005).
6. Turner, W.C., Ed., *Energy Management Handbook* 4th Ed., Fairmount Press: Lilburn, GA, 2001.

Facility Power Distribution Systems

Paul L. Villeneuve
University of Maine, Orono, Maine, U.S.

Abstract
Facility power distribution systems are very diverse. This diversity stems from a variety of items. Design code changes have resulted in evolving distribution system requirements. Furthermore, the owners and operators of the facilities have a varied set of concerns (financial, capacity, safety, etc.) that are to be incorporated into a power distribution system's design. Finally, a designer's experience often dictates the level of complexity that a design may entail. This entry attempts to address some of the basic concerns of the facility power distribution system designer. It provides a brief technical background as well as some straightforward recommendations.

INTRODUCTION

Electricity has been used in commercial and industrial facilities for over a century. Much of the manual labor that was routinely done has been replaced by relatively efficient motor-driven equipment. Furthermore, better monitoring and control has resulted in improved efficiency, higher yields, reduced emissions, and better working environments. It is fair to say that our ability to harness the power in electricity has been the most significant factor in improving our standard of living, enhancing our safety, and making our lives better.

The energy delivered by electricity has done many good things; however, that same energy can be destructive. Numerous fires, injuries, and fatalities have occurred due to the incorrect design of power systems or the failure of equipment. As a result, many design codes and standards governing power systems and equipment have been developed. These standards primarily focus on personnel and equipment safety. As a result, safety is often the most important consideration when designing or operating a facility power system. Further, the systems that are designed must be flexible enough to allow multiple operating conditions or plant configurations. Finally, power systems would not be viable unless they are economically feasible. As a result of all these considerations, it is the responsibility of the power system designer to ensure a safe system that provides for flexible equipment operation yet is economical to operate.

Because of increasing constraints and safety considerations, power systems continue to evolve. Until recently, on-site generation was not economically feasible for the majority of end users. Today, emerging technologies are allowing on-site generation which provide for additional sources of power, reducing energy costs, and enhancing efficiency. Electric utility providers have legitimate concerns regarding the effects that the on-site generation has on system reliability. As a result, these on-site generators create additional complexities in the design of facility power systems which the power system designer must factor in.

Keywords: Distribution; Power; Facility; Coordination; Fault analysis.

BASIC PRINCIPLES

Single- and Three-Phase Systems

Incoming utility power can be from a single-phase supply or a three-phase supply. A single-phase supply indicates that there is one independent phase capable of supplying electric power. However, most single-phase systems employ three separate conductors where current flow is permitted. These conductors originate at a single-phase utility step-down transformer. Two of the conductors are at the transformer's winding extremes. The third conductor is at the center of the transformer and is always tied to an earth ground. This type of configuration is called a center-tapped configuration. Single-phase systems were the original designs used in the early 1900s to provide power to all customers.

Three-phase systems were introduced to address the fact that large single-phase power requirements required large conductors. In a three-phase system, three independent phases are capable of supplying power. In most facility power distribution systems, there are four separate conductors where current flow is permitted; three-phase conductors and one neutral conductor. The neutral conductor is primarily used as a ground reference but can support current flow. In fact, three-phase systems with a neutral conductor can be used to provide phase-to-phase or phase-to-ground power. However, utilities require end users to balance their load so that each of the three-phases sees approximately the same load. This, in turn, limits the amount of neutral current flow.

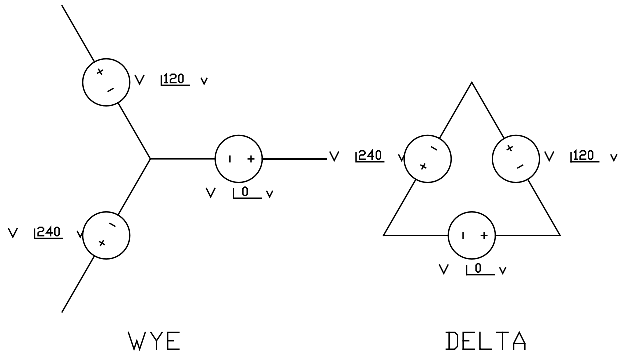


Fig. 1 Delta and wye configurations.

The neutral conductor is formed from the configuration of the utility's incoming three-phase transformer. The transformer can be configured in a delta or wye arrangement. The terms delta or wye come from the fact that the wye configuration looks like the letter Y and the delta configuration looks like the greek letter Δ. Fig. 1 shows how incoming power sources can be configured as a wye or a delta.

As shown in the figure, the center point of the wye connection can be physically accessed. This is the neutral point of the configuration. Note that there is no neutral point in the delta connection. The vast majority of three-phase distribution systems are tied together in a wye configuration.

With three-phase systems that are supplied by a wye configuration, there are two voltage magnitudes available. Looking at the figure, we see that there is a line to neutral voltage for each phase. This voltage is called the phase voltage or line-to-neutral voltage. To find the voltage magnitude across any two phases we apply Khirchoff's Voltage Law. Khirchoff's Voltage Law is based on the principle of conservation of energy and states that the sum of the voltages around a closed loop must equal zero. In a balanced system, the voltage magnitude across any two phases can be found from the magnitude of any phase voltage by multiplying the phase voltage by $\sqrt{3}$. Generally, loads are connected in a phase-to-phase configuration. This results in reduced cable sizes, due to the smaller levels of current demanded, and helps to ensure a balanced system.

Since single-phase power distribution systems have fairly straightforward design criteria, they are generally less complex. As a result, this article will emphasize the designs of three-phase power distribution systems.

Balanced and Unbalanced Systems

As we have already mentioned, the terms balanced and unbalanced describe the load that is on each phase relative to the other phases. In addition, these terms also describe the power supply. For a balanced system, all voltages must have identical magnitudes when measuring across any pair of phases. That is, the voltage vectors should have the same magnitude. Further, since the voltages are vectors, they must be evenly spaced by 120°. Refer to Fig. 1 to see how both delta- and wye-configured circuits are setup to be balanced.

Balanced loads are also preferred. Most three-phase loads are inherently balanced. For example, a three-phase motor has three independent sets of windings that are identical. Each of these windings is configured so that all three phases output the same mechanical power. As a result, the current delivered to each phase is approximately the same.

Unbalanced systems are those in which the supply is unbalanced and/or the loads are unbalanced. The vast majority of supplies are balanced. In addition, end users have no control over the supplies, so we will assume that if there is an unbalance, it is only because the loads are unbalanced. Loads can be connected to a wye-configured supply to access the phase-to-phase or phase-to-neutral voltage. In a delta configuration, only the phase-to-phase voltages are available. In either case, an unbalanced system occurs when all three phases do not have equal loading. To balance the load, simply reconnect the single-phase loads to the source so that they appear to carry the same load. Oftentimes, it is not possible to have a perfectly balanced system, but the power system designer should make all attempts to minimize the imbalance to within 10% of the loading on all phases.

In wye-configured systems, the unbalance results in the neutral voltage no longer being at 0 V. Since most wye-configured systems are grounded, a neutral current will flow through the ground. The higher the imbalance, the higher the neutral current flow will be. Transformers with primarily three-phase loads (most often medium voltage secondaries) have smaller neutral conductors than the phase conductors. Also, it is not unreasonable to use an uninsulated conductor to solidly tie the neutral of a three-phase system to a ground. Current through the neutral will cause a voltage to exist on the neutral, and improper operation may occur due to the uninsulated cable. Note that the grounding conductor is not the same as the neutral conductor for systems intending to supply single-phase loads. In the case of the system supplying single-phase loads, the neutral conductor must be insulated.

In delta-configured systems, the unbalance can be more troublesome. This is due to the fact that there is no neutral connection in which the unbalanced current can flow. The unbalanced current flow will actually travel within the delta causing the windings to heat up. This is an undesirable characteristic, as heating effects are occurring without any real work being accomplished. Furthermore, monitoring of unbalanced systems in a delta configuration is more difficult than in a wye configuration. The neutral connection of a wye configuration is generally readily accessible and monitored. Since the delta configuration has no neutral, a special configuration must be employed.

GENERAL SYSTEM DESIGN GUIDELINES

Grounding

Grounding is the most important factor when designing a facility's power distribution system. Good grounding results in high personnel safety, limits noise affects, and extends equipment life. Although grounding seems like a straightforward design consideration, there are a number of issues that make grounding somewhat of an art. In fact, there are numerous books and papers written on different grounding techniques and applications.

The key for grounding is to ensure safety of personnel. In ungrounded systems, an inadvertent ground will often go undetected. If personnel were to also inadvertently contact an ungrounded system with an inadvertent ground, they would provide a path to ground through their body and serious injury or death may result. By grounding the system, the first inadvertent ground is immediately detected and the system is usually taken offline automatically. This precludes personnel from any immediate danger. A very general rule of thumb to ensure proper grounding for personnel safety is to have numerous grounds. However, numerous grounds can result in multiple return paths and need to be evaluated, particularly when grounding the incoming service from the utility. Having numerous equipment grounds generally does not pose a problem, but they must be adequately sized so that faults can be cleared before the ground connections fail. National Fire Protection Agency (NFPA) Standard 70 (The National Electric Code) provides a table of required ground sizes based on the rating of the supply. Furthermore, the Institute of Electrical and Electronics Engineers (IEEE) provides Standard 142 (the Green Book), which contains many recommended grounding practices.

Numerous equipment grounds can introduce noise into the ground system. This is due to the fact that the ground system appears to be a closed loop permitting electromagnetic energy in the area to induce currents and voltages in the ground system. These induced currents and voltages create problems particularly for sensitive electronic equipment. To avoid induced currents and voltages in the ground system, use caution when closing the ground loop. Another option is to utilize an isolated ground for sensitive electronic equipment that is separate from the facility's power distribution system ground.

Grounding also provides protection against lightning phenomena. Much research has been done in the field to protect structures against lighting strikes. Detailed lightning protection is beyond the scope of this paper; however, interested readers may refer to NFPA 780 for recommendations.

Conductor Sizing

Cable and conductor sizing requires an understanding of the effects of current through a conductor. All materials will resist the flow of current through them. The term resistance describes the tendency to restrict to the current flow and is denoted in ohms (Ω). The higher the resistance, the harder the system must "work" to get current to flow. The work translates to power dissipation as heat. Clearly, a facility power distribution system's goal is low resistance to yield higher efficiencies.

The two most common conductors are copper and aluminum. These two are used due to their low resistivity (or high conductivity) and relatively low cost. Copper is more commonly used in insulated cables due to its lower resistance per unit area in addition to being more flexible. Aluminum conductors are more commonly used when the conductor is uninsulated as in the case of overhead cables on a utility pole line. In this case, the aluminum is reinforced with steel strands to increase its strength. Copper and aluminum are not always interchangeable. Due to a phenomenon called a galvanic reaction, copper to aluminum connections result in conductor corrosion. As a result, special terminators must be used when the two metals are to be interconnected.

Most heavier conductors utilize stranded conductors, i.e., they are made up of a number of smaller single strands. Stranding results in the improved flexibility of the conductor. Additionally, stranding also results in a lower overall AC resistance as opposed to a conductor made of a single strand with the same area. This is a result of a phenomenon called the skin effect. Extra flexible cable, sometimes called weld or locomotive cable, is used when specialty applications arise.

The amount of expected current to be delivered to a load helps to determine which cable size to use. The rule generally goes the higher the current, the larger the cable. In the United States, cables are sized using the American Wire Gauge (AWG). Typical power cable sizes in commercial or industrial facilities run from #12AWG to 750 thousands of circular mils (MCM). In addition to the current through the conductor, the ability to dissipate heat also dictates the conductor's size. For example, if many conductors are packed in close proximity to each other they will contribute to localized heating, increasing the ambient temperatures. In addition, conductors in air exhibit better cooling than those that are routed in raceway due to convection cooling. As a result, where the cable is routed must be taken into account when determining the required cable size.

For large loads, 750 MCM cable may not be sufficient. Options include utilizing a copper or aluminum bus as opposed to using cable or routing multiple conductors per phase. It is more common to use multiple conductors per phase due to the flexibility of the cable. There are a few concerns when using multiple conductors per phase. First, all the conductors must be the same guage and material (i.e., they must all be 500 MCM copper). Secondly, the corresponding number of conductors from the other phases must be routed in the same raceway. This is due to the fact the currents will be induced in the raceway unless the magnetic fields resulting from the current flow are cancelled out.

Additionally, the phase conductors have an insulation covering to prevent faults to ground. The heat generated in the conductor as a result of current flow must be dissipated through the insulation to the surrounding environment. This results in the conductor insulation being at a higher temperature than ambient. The heating effect results in reduced insulation life. As a result, manufacturers have developed various types of conductor insulation that can handle different temperatures. The conductors' insulation is typically rated for temperatures 60, 75, 90, or 105°C above an ambient temperature of 40°C. However, the power system designer must use caution when specifying insulation for a 90 or 105°C rise since many terminals of the end equipment are only rated for a 75°C increase.

Voltage Levels

A key design consideration in designing a facility power system is the voltage level employed at the facility. Newer industrial facility power distribution systems typically employ three-phase, 480 V incoming supplies while commercial facilities typically employ three-phase, 208 V or single-phase, 240/120 V. In either case, these systems are termed low voltage systems. In fact, voltage levels below 1000 V are considered low voltage systems. Some large industrial facilities may utilize medium voltage levels. A typical medium voltage used in industrial facilities is 4160 V, but may go as high as 13.8 kV. Higher medium voltage systems exist but are rarely found in industrial facilities. These higher voltages are typically found at generating facilities.

For industrial facilities, the items that dictate the necessary voltage levels are the output horsepower ratings of the motors in use at the facility. The greater the output ratings of the motors, the greater the current requirements are, which result in larger conductor sizes. Current industry practice is to specify motors up to about 3/4 HP as single-phase, 240 V motors, although single-phase, 240 V motors are available to about 10 HP. Motors that are greater than 3/4 HP, but less than either 250 or 350 HP are typically specified as three-phase, 480 V motors. Motors above 250 or 350 HP are usually specified as medium voltage motors.

SYSTEM PROTECTION

General Design Guidelines

Inevitably, all facility power distributions will experience faults. However, in a well designed system, equipment damage can be limited. Also, the location of the fault can be quickly located and repaired. Furthermore, isolation of some of the faults will allow the facility to continue to operate while the fault is being repaired.

In a poorly designed facility power distribution system, large portions of the facility will have to be taken off-line while a fault is repaired. Also, it may be very difficult to determine where the fault occurred. Furthermore, significant equipment damage is more likely in a poorly designed system, resulting in a long shutdown of the facility's power distribution system.

It is the responsibility of the power system's design engineer to determine the necessary system requirements. For example, a facility working with benign chemicals and relatively low temperatures may be willing to have a very simple power distribution system. On the other hand, a nuclear plant requires multiple layers of redundancy with complex switching schemes. This latter system would be very expensive, but the expense is required to ensure overall facility safety.

Power system protection has two separate concerns. The first is to detect a fault and the second is to isolate the fault. Fault detection depends upon the fault type. Isolation of the fault not only depends on the fault type but also on the severity of the fault.

Power System Faults

When we mention faults on a power system we are usually talking about short circuits. A short circuit occurs when a relatively low impedance path back to the source occurs. A low impedance path allows for current levels larger than the facility was designed for. The two most common types of short circuits are the single-line to ground fault and the three-phase fault. The single-line to ground fault is characterized by a single phase being subjected to a path to ground that is not anticipated. The three-phase fault is when all three phases are tied together with a low resistance.

The detection of faults is usually accomplished by seeing a larger than anticipated current in any phase or neutral of the conductor. The anticipated current can vary depending upon what is occurring in the system. For example, a motor starting and a transformer energizing result in very large but momentary currents, typically called inrush currents. This current is anticipated and the facility's power distribution system should be able to handle this current level without fault clearing. If these high levels of current are sustained, that is they are not momentary, then the protection system needs to react. Inrush levels of current exist for a few milliseconds for transformers or for a few milliseconds up to a minute depending upon the mechanical load connected to a motor. If the current is above these momentary levels, then a fault has likely occurred. Also, if the current is above the momentary levels for a longer than anticipated time, its likely that the motor has jammed. In either case, the primary difference between the jam and the fault condition is that the system reacts in a short time for sustained currents at the inrush level that exist longer than anticipated or trips instantaneously for currents above the inrush level. As a result, many facility distribution protection systems are equipped with devices that monitor and isolate systems under the instantaneous and short-time conditions.

In addition to the two conditions previously described, equipment failures may occur that are not immediately catastrophic but will be catastrophic to equipment if they are allowed to fester. However, if temporary overload operation of equipment that does not damage the equipment frequently occurs, the power distribution system must be able to identify when an overload has occurred versus when a fault has occurred. This operation allows for an above-normal current for a long time before tripping and is referred to as a long-time condition.

Another area of concern is that not all faults experience low-resistance faults. In the case of high-resistance faults, monitoring can be extremely challenging. However, most high-impedance faults will eventually switch to a low-impedance fault due to the destructive nature of the current. As a result, it is generally not a concern to have the additional detection devices for high-impedance faults.

Ground Fault Detection Methods

In most three-phase power distribution systems, neutral current is almost zero since the system is balanced. As a result, we can use the neutral current to detect minimal changes when an unanticipated unbalance occurs. This is typically denoted as ground fault protection. Ground fault protection can be accomplished directly by monitoring the current traveling through the neutral. Another option to monitor for a ground fault is to apply monitoring around all three of the phase conductors. If the current is balanced, no current is induced in the monitoring system. In an unbalanced situation, some residual current does not return through the three phases and is detected in the monitoring system. This is called a residual configuration. Ground fault protection is frequently used to detect when a single-line to ground fault occurs.

FAULT CURRENT

Fault Current Support

One key concept that is often misunderstood is that a sufficient fault current must exist to ensure mechanical device tripping. Whenever a fault occurs, voltage on the faulted phases tends to decrease. If the supply is weak, the voltage drop may be so significant that the fault current drops to a level below the protection of a system trip. In fact, a typical specification for on-site generation is to be able to provide 300% of the normal full load current for at least 10 s to allow a sufficient fault current to trip devices.

Fault Current Energy Sources and Levels

Due to the nature of faults, fault currents derive their energy not only from the power sources but also from rotating machines. Rotating machines operate on the principle of generating an internal voltage proportional to the mechanical load to limit the current flowing into the motor. This internal voltage is generated from the rotation of magnetic fields within the machine. When a fault occurs, the machine's terminal voltage typically goes to zero volts and the internal voltage source provides the fault current. In some machines, the fault current is only available for the first few electrical cycles (called a subtransient period). Other machines have several fields which allow for a fault current flow after the first few cycles. This is called the transient period and can last for up to 30 electrical cycles. Finally, machines that are operating as generators have external prime movers and voltage sources providing the necessary energy to provide fault current indefinitely. This period is called the steady-state.

All three currents can exist simultaneously. That is, the subtransient, transient, and steady-state fault currents exist immediately after the fault occurs. After two or three electrical cycles, the subtransient fault current decays away due to the rapid release of energy and only the transient and steady-state currents remain. After approximately 30 cycles, the transient fault current decays away leaving only the steady-state fault current. As a result, the magnitude of the current is greatest immediately after the fault has occurred and then decays to the constant steady-state level.

Fault Current Interrupting Devices

One of the properties associated with current flow is that a force is induced when current flows. The rule generally goes the greater the current, the greater the force. For proper protection, the system must be able to isolate faults by opening protective devices. Opening of these devices is accomplished with mechanical force. The mechanical force to open the device must be larger than the mechanical force that is induced from the fault current. As a result, protective devices are rated for a fault current rating as well as a continuous current rating. The continuous current rating is what the protective device can handle indefinitely without premature failure.

When a fault occurs, the protective device must be able to withstand the initial subtransient, plus transient, plus steady-state, currents. It is not required to interrupt this initial current surge as most protective devices cannot operate fast enough to detect the fault and then operate to interrupt the fault. In fact, many protective devices are rated for 3-cycle or 5-cycle operation which indicates the amount of time it takes for the device to operate.

For fuse-type protective devices, fault interruption is typically accomplished by a metallic element that melts due to the excessive heating caused by the large current flow. As a result, fuses can operate extremely fast. In fact, fuses are often used to provide a current limiting capability. The practice of current limiting introduces a resistance in line (in series) with the fault, thereby limiting its magnitude. Typically, current limiting fuses operate to limit the fault current within one-fourth of the electrical cycle. Current limiting fuses are often used in series with other protective devices that are not rated for as large a fault current interrupting capability as required. This frequently used technique is called series rating. However, NFPA 70 does not permit series rating unless the two protective devices in series have been properly tested and are UL listed for series rating. Current limiting fuses can have fault current interrupting capabilities to a fault current level of 200,000 Amps. Frequently, power system engineers utilize current limiting fuses to reduce fault current levels to more manageable levels. Another advantage of fuses is that they are relatively inexpensive.

There are several major disadvantages to utilizing fuses as protective devices. One disadvantage is that in the event that only one or two of the phases experiences a fault, only one or two of the fuses will operate. When this occurs, the supply source becomes unbalanced and the load is said to be "single-phasing." This situation can lead to equipment damage. As a result, fuse manufacturers have built devices that detect when one fuse has operated and then causes the remaining fuses to operate. Another disadvantage is that when a fuse operates, it can no longer be used. Although repeated faults are not expected under normal operation, repeated faults may occur particularly when a facility startup is occurring. This would require a large quantity of spare fuses which may not be desirable.

Another disadvantage is that fuse characteristics cannot be changed for a specific fuse. As a result, if different characteristics are desired, new fuses must be installed. In an industrial facility, it would not be uncommon to have many different protection requirements thereby requiring a large number of different fuses.

A fourth disadvantage, and the most significant, is that there is no way to remotely operate the fuses. That is, an operator must be physically at the switch that contains the fuses to either open or close the circuit supplying the load. Remote operation is frequently referred to as electrical operation as there is an electrical operator that interfaces with the device to open and close circuits. For loads that are always operating, remote operation isn't usually a requirement. However, many facility loads are cycled on and off and utilizing human operators to operate the protective devices is not practical.

Circuit Breakers and Contactors

Circuit breakers and contractors are two common devices that are utilized to allow remote operation. Not all circuit breakers are required to have remote operation capabilities. In fact, the design of circuit breakers is such that they will only operate sporadically. Usually, their operation is the result of a fault where special monitoring is required. Some of these breakers are only equipped with a shunt trip capability which only allows for breaker tripping. Contactors, on the other hand, are designed for repeated opening and closing but are not designed to be able to interrupt large fault currents. In fact, contactors usually only occur in series with fuses that will actually interrupt the large fault currents. The operation of the contactors during large fault currents is inhibited.

FAULT CURRENT COORDINATION

In many facility power distribution systems, it is very common to have a single main power distribution center. Smaller distribution systems that are fed from the main power distribution center are located throughout the facility for the purpose of concentrated power distribution. To limit the effect on the overall facility, a power system's designer should configure the system so that fault currents result in the minimum amount of service disruption. In order to accomplish this, upstream protective devices must coordinate their tripping characteristics with downstream devices. This is called protective device coordination and is accomplished by evaluating-time versus-current characteristic curves. A simple coordination curve is show in Fig. 2.

As shown on the curve, the *X*-axis is in amperes and the *Y*-axis is in seconds. For proper device coordination, the device closest to the fault should trip first. If the fault is upstream then the next closest device should trip. As a result, the trip curves for each protective device must be configured so that the device furthest downstream trips first and then progresses to the largest device. Relaying this to the curve requires that the trip curves progress from left to right for the farthest downstream device to the highest-level device.

In addition to protective device trip characteristics, coordination curves also contain motor starting characteristics and transformer inrush currents. These are shown so that devices are allowed to startup and run normally. In addition, cable melting times are shown. This is done to ensure that cables that supply current to the fault are not significantly damaged.

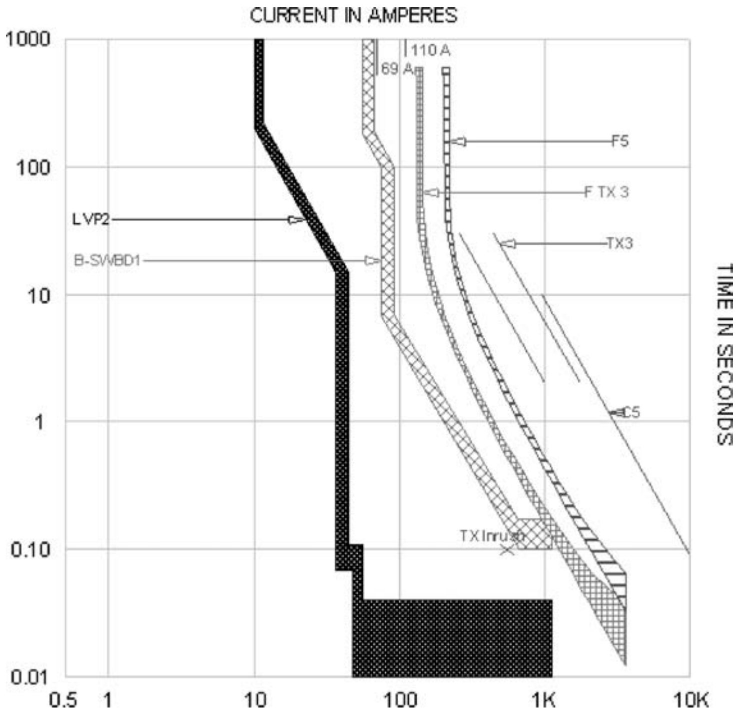


Fig. 2 Coordination curve.

OTHER FACILITY CONSIDERATIONS

Transformer Ratings

Specifying transformer parameters is another key concern for power system designers. The most important parameter is to specify the correct turns ratios so that proper levels are achieved. Typically, the designer will include transformer taps that can be adjusted when the transformer is de-energized. These taps allow periodic changes in supply voltages or loads and still provide approximate nominal voltages.

Another important parameter to specify is the continuous power rating of the transformer. The power system designer has to know the expected load of the facility as well as the anticipated growth of the facility load. If spurious conditions result in large load increases, transformers can be equipped with fan banks to increase their nominal ratings. In addition, the designer can specify a nominal power rating at a lower temperature rise than what the transformer insulation can handle. Then, when an overload is required, the transformer can continue to operate at a higher temperature rise without the insulation being damaged.

The other major concern for the power system designer is specifying the transformer characteristic impedance (normally called %Z). This value helps to determine the efficiency of the transformer. It also affects the fault current through a system. Typical transformer %Z values range from 2.25 to 5.75%. Transformer characteristic impedances are of large concern for facilities with generation capability. This is due to the fact that utilities require voltage support and may require the facility to reduce their generation to meet interconnection requirements. The transformer characteristic impedance consumes the magnetic energy that supports the voltage. Furthermore, large characteristic impedances cause large voltage drops in the system when high current demands exist. This may result in the misoperation of certain equipment due to the low voltage.

Direct Current Systems

Direct current systems are used in some facilities for critical or life-safety systems. Direct current power is readily available from batteries. As a result, when incoming utility power is lost, the batteries can support essential systems for a limited time. Direct current systems are so vital to the plant that they are permitted to be ungrounded to ensure their continued operation. However, even though these systems are ungrounded, ground detection systems are usually provided to indicate that a ground has occurred. This perpetuates the concept of maintaining personnel safety.

Variable Frequency or Variable Speed Drives

In older processes, many inefficient designs existed. This was acceptable since energy costs were relatively low and there was little equipment available to enhance system efficiency. However, the proliferation of variable frequency or variable speed drives has enhanced system efficiency. These drives operate by either adjusting the frequency or voltage of the power supply source. A major disadvantage of these drives is that they tend to create a non-sinusoidal output waveform. This results in harmonics that can cause equipment damage. Newer drive topologies utilize a higher number of pulses to create a more sinusoidal output waveform. Some drives still require isolation transformers to keep the harmonics from traveling back to the supply and negatively affecting the system.

Power Factor

Industrial facilities frequently have a poor power factor. The power factor determines the ratio of real power consumed to the total supplied power. A portion of the supplied power is considered "imaginary" in that it does no real work but is necessary for proper device operation. Generally, utilities require industrial facilities to have at least a 0.9 lagging power factor. The lagging term simply indicates that the industrial facility is receiving imaginary power, as opposed to supplying imaginary power (leading). If the facility's power factor is below 0.9 lagging, then the utility will charge a power factor penalty to the industrial facility. In the case where a poor power factor exists at a facility, power factor correction capacitors are specified for a facility. In fact, a common option when purchasing contactors for motor control is to include power factor correction capacitors since motors are the most frequent cause of poor power factors.

CONCLUSION

This paper has discussed the key items for consideration when designing and operating a facility's power system. It has attempted to give the reader a brief technical background of each of these items. while emphasizing the importance that safety plays in the design of a system. Although continued operation is desired, it is not more important than personnel safety. By taking into consideration the concepts discussed in this article, the reader should have a basic understanding of the items involved in designing a facility's power system.

ACKNOWLEDGMENTS

The author thanks Dr. Scott C. Dunning, Mr. Benjamin Standish, and Mr Thomas Osborne for their helpful review comments.

BIBLIOGRAPHY

1. Chapman, S.J. *Electric Machinery and Power System Fundamentals*; McGraw Hill: New York, 2002.
2. Villeneuve, P.L. Concerns generated by islanding. IEEE Power Energy Mag. **2004**, *May/June*.
3. *Cutler-Hammer 2003 Consulting Specifier's Guide*, CA 08104001E.
4. 2005 NFPA 70, The National Electric Code.
5. IEEE Standard 142, The Green Book.
6. NFPA 780, Standard for the Installation of Lightning Protection Systems.

Federal Energy Management Program (FEMP): Operations and Maintenance Best Practices Guide (O&M BPG)☆

Gregory P. Sullivan
Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.

W. David Hunt
Pacific Northwest National Laboratory, Washington, D.C., U.S.A.

Ab Ream
U.S. Department of Energy, Washington, D.C., U.S.A.

Abstract

The Federal Energy Management Program's (FEMP'S) Operations and Maintenance Best Practices Guide (O&M BPG) highlights O&M programs targeting energy efficiency, which are estimated to save 5%–20% on energy bills without a significant capital investment. Depending on the federal site, these savings can represent thousands to hundreds of thousands of dollars each year, and many can be achieved with minimal cash outlays. In addition to energy/resource savings, a well-run O&M program will:

- Increase the safety of all staff because properly maintained equipment is safer equipment.
- Ensure the comfort, health, and safety of building occupants through properly functioning equipment, providing a healthy indoor environment.
- Confirm the design life expectancy of equipment.
- Facilitate compliance with federal legislation such as the *clean air act* and the *clean water act*.

The focus of this guide is to provide the federal O&M/energy manager and practitioner with information and actions aimed at achieving these savings and benefits. The O&M BPG was developed under the direction of the Department of Energy's Federal Energy Management Program.

GUIDE DESIGN AND LAYOUT

The guide currently consists of 11 chapters. Chapter 1 provides an introduction and an overview. Chapter 2 provides the rationale for "Why O&M?" Chapter 3 discusses O&M management issues and their importance. Chapter 4 examines computerized maintenance management systems (CMMS) and their role in an effective O&M program. Chapter 5 looks at the different types of maintenance programs and definitions. Chapter 6 focuses on maintenance technologies, particularly the most accepted predictive technologies. Chapter 7 discusses the process of building commissioning, particularly as it applies to existing buildings. Chapter 8 highlights the importance of metering and its role in O&M. Chapter 9 explores O&M procedures for the predominant equipment found at most federal facilities. Chapter 10 describes some of the promising O&M technologies and tools on the horizon to increase O&M efficiency, and Chapter 11 provides 10 steps to initiating an *operational efficiency* program.

The O&M environment is in a constant state of evolution, and the technologies and vocabularies are ever-expanding. Therefore, the guide contains a glossary of terms in Appendix A. Appendix B provides a list of federal contacts for training and assistance. Appendix C includes a list of organizations and trade groups that have interest or are related to O&M. And finally, Appendix D is a form that can be used to submit suggestions or revisions to the guide.

The goal of this article is to provide an overview of the O&M BPG—how and why it was developed—and to present in highlight form some of its sections. Each highlighted section will be referenced in parentheses in this article's section title.

INTRODUCTION (CHAPTER 1)

Effective O&M is one of the most cost-effective methods for ensuring reliability, safety, and energy efficiency. Inadequate maintenance of energy using systems is a major cause of energy waste in both the federal government and the private sector. Energy losses from steam, water and air leaks, uninsulated lines, maladjusted

☆This entry originally appeared as "FEMP's O&M Best Practices Guide–A Guide to Achieving Operational Efficiency" in *Strategic Planning for Energy and the Environment*, Vol. 23, No. 4, 2004. Originally written by Gregory P. Sullivan, Aldo P. Melendez and Ray Pugh, updated by Gregory P. Sullivan, Dave Hunt, and Ab Ream. Reprinted with permission from AEE/Fairmont Press.

Keywords: Operations; Maintenance; Efficiency; Metering; Commissioning; Preventive; Predictive; Reactive.

or inoperable controls, and other losses from poor maintenance are often considerable. Good maintenance practices can generate substantial energy savings and should be considered a resource. Moreover, improvements to facility maintenance programs can often be accomplished immediately and at a relatively low cost.

The purpose of the O&M BPG is to provide the operation and maintenance/energy manager and practitioner with useful information about O&M management, technologies, and cost-reduction approaches. To make this guide useful and to reflect the facility manager's needs and concerns, the authors met with O&M and energy managers via federal energy management program (FEMP) workshops. In addition, the authors conducted extensive literature searches and contacted numerous vendors and industry experts. The information and case studies that appear in the guide resulted from these activities.

It needs to be stated at the outset that the guide is designed to provide information on effective O&M as it applies to systems and equipment typically found at federal facilities. The guide "is not" designed to provide the reader with step-by-step procedures for performing O&M on any specific piece of equipment. Rather, the guide first directs the user to the manufacturer's specifications and recommendations. In no way should the recommendations in the guide be used in place of manufacturer's recommendations. The recommendations in the guide are designed to supplement those of the manufacturer, or, as is all too often the case, provide guidance for systems and equipment for which all technical documentation has been lost.

WHY O&M? (CHAPTER 2)

Definitions

Operations and maintenance. Decisions and actions regarding the control and upkeep of property and equipment is inclusive but not limited to the following: actions focused on scheduling, procedures, and work/systems control and optimization, and performance of routine, preventive, predictive, scheduled and unscheduled actions aimed at preventing equipment failure or decline with the goal of increasing efficiency, reliability, and safety.

Operational efficiency. It represents the life cycle cost-effective mix of preventive, predictive, and reliability centered maintenance technologies, coupled with equipment calibration, tracking, and computerized maintenance management capabilities, all targeting reliability, safety, occupant comfort, and system efficiency.

O&M Potential, Energy Savings, and Beyond

It has been estimated that O&M programs targeting energy efficiency can save between 5 and 20% on energy bills without a significant capital investment.[1] From small to large sites, these savings can represent thousands to hundreds of thousands of dollars each year, and many can be achieved with minimal cash outlays. Beyond the potential for significant cost and energy/resource savings, an O&M program operating at its peak "operational efficiency" has other important implications:

- A well-functioning O&M program is a safe O&M program. Equipment is maintained properly, mitigating any potential hazard arising from deferred maintenance.
- In most federal buildings, the O&M staff are responsible for not only the comfort, but also the health and safety of the occupants. Indoor air quality (IAQ) issues within these buildings are of increasing productivity (and legal) concern. Proper O&M reduces the risks associated with development of dangerous and costly IAQ situations.
- Properly performed O&M helps ensure that the design life expectancy of equipment can be achieved, and in some cases, exceeded. Conversely, the costs associated with early equipment failure are usually not budgeted for, and often come at the expense of other planned activities.
- An effective O&M program aids in a facility's compliance with federal legislation, such as the *clean air act* and the *clean water act*.
- A well functioning O&M program is not always reactive (answering complaints); rather, it is proactive in its response, and corrects situations before they become problems. This model minimizes callbacks and keeps occupants satisfied while allowing more time for scheduled maintenance.

O&M MANAGEMENT (CHAPTER 3)

O&M management is a critical component of the overall program. The management function should bind the distinct parts of the program into a cohesive entity. From our experience, the overall program should contain five very distinct functions making up the organization[2]: Operations, Maintenance, Engineering, Training, and Administration—OMETA.

Beyond establishing and facilitating the OMETA links, O&M managers have the responsibility of interfacing with other department managers and making their case against ever-shrinking budgets. Their roles also include project implementation functions, as well as the need to maintain the program and its goals.

Developing the Structure

Five well-defined elements of an effective O&M program include those presented above in the OMETA concept.[2]

While these elements (operations, maintenance, engineering, training, and administration) form the basis for a solid O&M organization, the key lies in the well-defined functions each brings and the linkages between organizations. A subset of the roles and responsibilities for each of the elements is presented in the guide; further information is found in Meador.[2]

Obtain Management Support

Federal O&M managers need to obtain full support from their management structure to carry out an effective maintenance program. A good way to start is by establishing a written maintenance plan and obtaining upper management approval. Such a management-supported program is very important because it allows necessary activities to be scheduled with the same priority as other management actions. Approaching O&M by equating it with increased productivity, energy efficiency, safety, and customer satisfaction are ways to gain management attention and support.

Measuring the Quality of your O&M Program

Traditional thinking in the O&M field focused on a single metric, reliability, for program evaluation. Every O&M manager wants a reliable facility; however, this metric alone is not enough to evaluate or build a successful O&M program.

Beyond reliability, O&M managers need to be responsible for controlling costs, evaluating and implementing new technologies, tracking and reporting on health and safety issues, and expanding their program. To support these activities, the O&M manager must be aware of various indicators that can be used to measure the quality or effectiveness of the O&M program. Not only are these metrics useful in assessing effectiveness, but they are also useful in the justification of costs for equipment purchases, program modifications, and staff hiring.

Below are a number of metrics that can be used to evaluate an O&M program. Not all of these metrics can be used in all situations; however, a program should use as many metrics as possible to better define deficiencies and, most importantly, publicize successes. The O&M BPG describes each of these metrics in detail.

- Capacity factor
- Work orders generated/closed
- Backlog of corrective maintenance
- Safety record
- Energy use
- Inventory
- Overtime worked
- Environmental record
- Absentee rate
- Staff turnover.

Selling O&M to Management

To successfully interest the management in O&M activities, O&M managers need to be fluent in the language spoken by management. Projects and proposals brought forth to management need to stand on their own merits and be competitive with other funding requests. While evaluation criteria may differ, generally some level of economic criteria will be used. O&M managers need to have a working knowledge of economic metrics such as:

- *Simple payback*—The ratio of total installed cost to first-year savings.
- *Return on investment*—The ratio of the income or savings generated to the overall investment.
- *Net present value*—The present worth of future cash flows minus the initial cost of the project.
- *Life-cycle cost*—The present worth of all costs associated with a project.

Program Implementation

Developing or enhancing an O&M program requires patience and persistence. Guidelines for initiating a new O&M program will vary by agency and management situation; however, some steps to consider are presented below:

- *Start small*—Choose a project that is manageable and can be completed in a short period of time, 6 months to 1 year.
- *Select troubled equipment*—Choose a project that has visibility because of a problematic history.
- *Minimize risk*—Choose a project that will provide immediate and positive results. This project needs to be successful, and therefore, the risk of failure should be minimal.
- *Keep accurate records*—This project needs to stand on its own merits. Accurate, if not conservative, records are critical to compare before and after results.
- *Tout the success*—When you are successful, this needs to be shared with those involved and with management. Consider developing a "wall of accomplishment" and locate it in a place where management will take notice.
- *Build off this success*—Generate the success, acknowledge those involved, publicize it, and then request more money/time/resources for the next project.

O&M Contracting

Approximately 40 percent of all non-residential buildings contract maintenance service for heating, ventilation, and air conditioning (HVAC) equipment.[3] Discussions with federal building managers and organizations indicate that this value is significantly higher in the federal sector, and the trend is toward increased reliance on contracted services. The O&M BPG explores this trend further and offers guidance on O&M contracting.

COMPUTERIZED MAINTENANCE MANAGEMENT SYSTEMS (CHAPTER 4)

A computerized maintenance management system (CMMS) is a type of management software that performs functions in support of management and tracking of O&M activities. CMMS systems automate most of the logistical functions performed by maintenance staff and management. CMMS systems come with many options and have many advantages over manual maintenance tracking systems. The O&M BPG presents the major capabilities, benefits, and potential pitfalls of CMMS.

TYPES OF MAINTENANCE PROGRAMS (CHAPTER 5)

What is maintenance and why is it performed? Past and current maintenance practices in both the private and government sectors would imply that maintenance is the actions associated with equipment repair after it is broken. The dictionary defines maintenance as follows: "The work of keeping something in proper condition; upkeep." This would imply that maintenance is actions taken to prevent a device or component from failing or to repair normal equipment degradation experienced with the operation of the device to keep it in proper working order. Unfortunately, data obtained in many studies over the past decade indicate that most private and government facilities do not expend the necessary resources to maintain equipment in proper working order. Rather, they wait for equipment failure to occur and then take whatever actions are necessary to repair or replace the equipment. Nothing lasts forever, and all equipment has some predefined life expectancy or operational life. For example, equipment may be designed to operate at full design load for 5000 h and may be designed to go through 15,000 start and stop cycles.

The design life of most equipment requires periodic maintenance. Belts need adjustment, alignment needs to be maintained, proper lubrication on rotating equipment is required, and so on. In some cases, certain components need replacement (e.g., a wheel bearing on a motor vehicle) to ensure the main piece of equipment (in this case a car) lasts for its design life. Anytime maintenance activities intended by the equipment's designer are not performed, we shorten the operating life of the equipment. But what options do we have? Over the last 30 years, different approaches to how maintenance can be performed to ensure, equipment reaches or exceeds its design life, have been developed in the United States. In addition to waiting for a piece of equipment to fail (reactive maintenance), we can utilize preventive maintenance, predictive maintenance, or reliability centered maintenance.

Chapter 5 of the O&M BPG provides a detailed description of each major maintenance program type (reactive, preventive, predictive, and reliability centered), including each program's advantages and disadvantages.

PREDICTIVE MAINTENANCE TECHNOLOGIES (CHAPTER 6)

Predictive maintenance attempts to detect the onset of a degradation mechanism with the goal of correcting that degradation prior to significant deterioration in the component or equipment.

The diagnostic capabilities of predictive maintenance technologies have increased in recent years with advances in sensor technologies. These advances, breakthroughs in component sensitivities, size reductions, and most importantly, costs have opened up an entirely new area of diagnostics to the O&M practitioner.

As with the introduction of any new technology, proper application and training is of critical importance. This is particularly true in the field of predictive maintenance technology, which has become increasingly sophisticated and technology driven. Most industry experts would agree (as well as most reputable equipment vendors) that this equipment should not be purchased for in-house use if there is not a serious commitment to proper implementation, operator training, and equipment upkeep. If such a commitment cannot be made, a site is well advised to seek other methods of program implementation,a preferable option may be to contract these services to an outside vendor and rely on their equipment and expertise.

Chapter 6 presents a detailed description and applications for predictive technologies including: thermography, oil analysis, ultrasonic analysis, vibration analysis, motor analysis, and performance trending.

COMMISSIONING EXISTING BUILDINGS (CHAPTER 7)

Commissioning of existing buildings is quickly becoming one of the most important topics in the building management arena. In general, commissioning is the process of ensuring that a building performs according to its design intent and its owner's and occupants' needs.[4] While additional research is needed to further pinpoint the costs and resulting benefits of commissioning new and existing buildings, numerous case studies have demonstrated resulting O&M-related energy efficiency improvements on the order of 5%–30% covering a wide range of building uses. Resulting simple payback periods are typically less than 2 years and often less than 0.5 years.

Definitions

There are a number of commissioning approaches that can be applied to building mechanical/electrical equipment and systems.

New building commissioning: New building commissioning (Cx) is a means to ensure through design reviews, functional testing, system documentation, and operator training that systems and equipment in new buildings are operating properly.

Re-commissioning: Re-commissioning (RCx), which is sometimes referred to as "retro-commissioning," is the practice of commissioning existing buildings—testing and adjusting the building systems to perform to satisfy the original design intent and/or optimize the systems to satisfy current operational needs. RCx relies on building and equipment documentation, along with functional testing to optimize performance.

Continual commissioning: Continual commissioning refers to a commissioning approach that is integrated into a facility's standard O&M program. As such, activities in support of the continual commissioning effort are completed on a regular basis, compared to recommissioning approaches that tend to be distinct events. By definition, continual commissioning works to ensure more stable building operations over time than the re-commissioning approaches.

Value re-commissioning: Value re-commissioning (VRCx) is the lowest cost option that focuses on the most common opportunities, ideally incorporating them into daily operating procedures. VRCx is the least comprehensive and requires the least specialized skillset. VRCx concentrates on the most common opportunities that typically carry the shortest payback periods. Therefore, VRCx is best applied in buildings where resources for structured re-commissioning or continual commissioning programs are not available. In addition to realizing highly cost-effective energy savings, tracking benefits (e.g., energy savings, cost savings, reduced occupant complaints) of VRCx activities can be helpful in developing justifications for funding requests of the more robust commissioning approaches.

Typical Findings from Existing Building Commissioning

Many case studies of existing building commissioning efforts have been published over the years. A review of case studies for multiple buildings published by Portland Energy Conservation, Inc. (PECI),[5] Texas A&M University,[6] Proceedings from National Building Commissioning Conferences, and FEMP Assessments of Load and Energy Reduction Techniques (ALERT) is useful in identifying measures most typically available in commercial building spaces. The most frequently cited measures/opportunities are:

- *Adjust reset temperatures and temperature settings*—settings are often adjusted over time based on personal preferences, to compensate for inadequate system operation, or to achieve energy savings. In addition, sensors require periodic recalibration.
- *Staging/sequencing of boilers, chillers, and air handling units*—equipment should be operated in the most efficient combination of chillers, boilers, and fans at varying load conditions.
- *Adjust and repair dampers and economizers*—malfunctioning or poorly tuned dampers (including seals, actuators, and linkages) and economizers result in (1) increased supply air fan energy in the closed position or require additional air heating and cooling when opened too much, (2) undesired building operating conditions due to lack of outside air, and (3) premature equipment degradation and replacement.
- *Modify control strategies for standard hours of operation*—motors, pumps, fans, and air handlers often operate on a 24/7 schedule even though not required by either the building tenants or the building operating plan.
- *Eliminate simultaneous heating and cooling*—heating and cooling systems for the same space can compete against each other due to improper set points.
- *Air and water distribution balancing and adjustments*—systems require rebalancing due to drift and changing building/workspace mission and/or tenant requirements.
- *Verify controls and control sequencing*—this includes enabling and re-enabling automatic controls for setpoints, weekends, and holidays. Verify that overrides are released.

The Commissioning Process

A four-step process for existing building commissioning is often recommended.[7]

Step 1: planning. The planning step includes developing and agreeing on the overall commissioning objectives and strategies, assembling the project team, and compiling and perusing building and equipment documentation. Examples of objectives could be a desire to optimize building operations to reduce operating costs, address complaints from occupants regarding air quality or comforts, create a model facility, and improve facility O&M including reducing emergency trouble calls. Regarding the commissioning team formation, considerations in forming the team could include contracted or in-house staff, level of effort required, desired and necessary qualifications, availability and use of resident knowledge, and available funding resources.

Step 2: Investigation. During this step the site assessment is completed, monitoring and functional test plans are developed and executed, test results are analyzed, a master list of deficiencies is compiled, and recommendations for improvements, including estimates of energy and cost savings, are generated and presented for consideration.

Step 3: Implementation. Accepted recommendations from the investigation step are put into place in the implementation step. Actions include making repairs and improvements, retesting and re-monitoring for results, fine-tuning improvements as needed, and revising estimated energy and cost savings.

Step 4: Hand-off and integration. Final documentation of the commissioning effort describing the process, individuals, systems information, and actions taken is developed in this step. Also developed is a plan for future commissioning efforts. Items addressed by the commissioning plan should include recommended procedures for specific building equipment, frequency of testing, analysis of results, periodic reporting, identification of key players, and budget requirements.

METERING FOR OPERATIONS AND MAINTENANCE (CHAPTER 8)

The Importance of Metering

Metering and sub-metering of energy and resource use is a critical component of a comprehensive O&M program. Metering for O&M and energy efficiency refers to the measurement of quantities of energy delivered (e.g., kilowatt-hours of electricity, cubic feet of natural gas, pounds of steam). Metering may also involve identifying times-of-use for the various energy sources, the instantaneous demand for energy, as well as identify energy use for a collection of buildings, individual buildings, rooms, or specific equipment (e.g., a boiler, chiller, or motor). With metering, you can:

- Monitor existing utility usage
- Verify utility bills
- Identify the best utility rate plans
- Measure, verify, and optimize equipment performance
- Isolate energy use and cost
- Measure, not estimate, tenant energy use
- Diagnose equipment and system operations
- Manage energy use.

Metering Applications

The uses for metered data vary from site-to-site and while not all sites have the same uses, some of the more common applications are presented below.[8]

- *Data recording*. Advanced meters can duplicate the conventional metering function of recording total consumption, plus offer enhanced functions such as time-of-use, peak demand, load survey, and power outage recording. For electric metering, advanced meters may also include recording of other electric characteristics, such as voltage, current, and power factor.
- *Total consumption*. This is the most basic data recording function, which duplicates the standard kilowatt-hour of electricity (kWh), hundred cubic feet volume (CCF) of gas, or gallons (gal) of water consumed between meter readings.
- *Time-of-use metering*. Different rates can be charged for on-peak and off-peak time periods by accumulating the total consumption during operator-defined time windows. The time windows may vary during both time of day and weekday/weekend/holiday.
- *Peak demand metering*. Billing of many larger commercial and industrial customers is based on total consumption and the highest 15-, 30-, or 60-minute demand during the billing period. The peak demand may be reported as a single highest value, highest four values, or highest value during each hour (all peak demand values must be accompanied by an associated time stamp).
- *Load survey (profile or time-series data)*. Energy consumption and conservation impact studies, as well as more complex analysis of system loading, require more detailed demand data. A load survey provides periodic consumption or demand data (in time increments of 1, 5, 15, 30, or 60 min).
- *Monitoring and control*. A two-way communication link between a central station and customer site provides the opportunity for integrating some other utility functions into the metering functions. Meters can be programmed to detect and report by exception (e.g., report only when a fault is detected) for power outage, leak detection, and tamper detection. The meter can also dispatch control functions, such as remote service disconnect/reconnect, demand-side management (DSM) load control, and load scheduling.
- *Load control*. Load control includes DSM control functions such as air conditioner and water heater load-shedding. The DSM load control could be triggered by a fixed algorithm operating independently or real-time central station control.
- *Load scheduling*. This includes scheduled start and stop of equipment to minimize or shift load to take maximum advantage of the demand and time-of-use billing rate structures.
- *Leak detection*. Continuous monitoring of gas or water usage can be used to detect leaks.

Steps in Meter Planning

The development of a federal metering program is highly dependant on a site's needs, existing metering equipment, and available infrastructure. When it comes to metering, *one size does not fit all*. Below are some very general guidelines identifying the steps and actions necessary for a quality metering program. These guidelines summarize information found in AEC;[9] EPRI;[10] and Sydlowski,[8] where more detailed information can be found.

- Formalize objectives and goals of metering program
  - —Identify and confirm goals of stakeholders/users
  - —Prioritize goals as near-term, mid-term, and long-term
  - —Formalize necessary/expected outcomes
- Develop program structure. Identify data needs, equipment needs, analysis methodologies, and responsible staff.
  - —Develop data and analysis needs based on necessary outcomes
  - —Develop equipment needs based on data needs
  - —Take advantage of existing infrastructure
  - —Identify responsible staff, preferably a metering "champion"
- Develop criteria for evaluating metering costs, benefits, and impacts to existing systems, infrastructure, and staff.
  - —Determine relative economics of proposal
  - —Justify with cost/benefit, return on investment, or payback metric
- Develop a prioritized implementation plan targeting manageable successes
  - —Screen opportunities based on success potential
  - —Start small/manageable—build off success
- Develop a sustainable plan targeting use, updates, calibration, maintenance, and program reinvestment.
  - —Maintain your investment
  - —Make this success visible
  - —Plan for future implementation/reinvestment.

O&M IDEAS FOR MAJOR EQUIPMENT TYPES (CHAPTER 9)

At the heart of all O&M lies the equipment. Across the federal sector, this equipment varies greatly in age, size, type, model, fuel used, condition, etc. While it is well beyond the scope of this guide to study all equipment types, we tried to focus our efforts on the more common types prevalent in the federal sector. The objectives of this chapter in the guide are:

- Present general equipment descriptions and operating principles for the major equipment types.
- Discuss the key maintenance components of that equipment.
- Highlight important safety issues.
- Point out cost and efficiency issues.
- Provide recommended general O&M activities in the form of checklists.
- Where possible, provide case studies.

The major equipment types covered in Chapter 9 include boilers, steam traps, chillers, cooling towers, energy management/building automation systems, pumps, fans, motors, air compressors, and lighting. At the end of each section in Chapter 9, a checklist of suggested O&M activities is provided. These checklists are not presented to replace activities specifically recommended by equipment vendors or manufacturers. In most cases, these checklists represent industry standard best practices for the given equipment. They are presented here to supplement existing O&M procedures, or to serve as reminders of activities that should be taking place.

O&M FRONTIERS (CHAPTER 10)

As old a topic as O&M is, there are a number of new technologies and tools targeting the increased efficiency of O&M. As with most new technology introduction, these tools are in various stages of commercialization; for up-to-date information on each tool contact information is provided. This chapter serves to highlight some of the more promising technologies targeting improved O&M and efficiency.

TEN STEPS TO *OPERATIONAL EFFICIENCY* (CHAPTER 11)

As defined, operational efficiency is the life-cycle, cost-effective mix of preventive, predictive, and reliability-centered maintenance technologies, coupled with equipment calibration, tracking, and computerized maintenance management capabilities, all targeting reliability, safety, occupant comfort, and system efficiency. Chapter 11 presents 10 simple steps to begin the path toward improved O&M and ultimately *operational efficiency*.

APPENDIXES

Four appendixes are provided in the O&M BPG. These are: Appendix A, a glossary of common terms; Appendix B, the FEMP staff contact list; Appendix C, a variety of O&M resources including relevant trade organizations and web sites; and Appendix D, a form to submit offering suggestions for the next version of the O&M BPG.

CONCLUSIONS

As FEMP's O&M program has matured, the O&M BPG has provided valuable guidance to federal building managers, O&M program managers, and building operations staffs. This guidance provides a starting point for establishing clear objectives and understanding benefits. It also can be used to (1) establish an effective long-range plan that involves all O&M-related staff functions, (2) measure existing program performance, (3) review and upgrade existing practices, and (4) plan for the future. This guide continues to assist federal building managers in realizing significant cost-effective energy savings and improved occupant satisfaction.

For More Information

To obtain information on the FEMP Operations and Maintenance Best Practices Guide, visit the O&M Program section on the FEMP Web Site at http://www.eere.energy.gov/femp/operations_maintenance/.

REFERENCES

1. PECI *Operations and Maintenance Assessments*; Prepared by Portland Energy Conservation, Inc. for the U.S. Environmental Protection Agency and U.S. Department of Energy: Washington, DC, 1999.
2. Meador, R.J. *Maintaining the Solution to Operations and Maintenance Efficiency Improvement*; World Energy Engineering Congress: Atlanta, GA, 1995.
3. PECI. *Energy Management Systems: A Practical Guide*; Portland Energy Conservation, Inc.: Portland, OR, 1997.
4. Gregerson, J. *Commissioning Existing Buildings*; E Source: Boulder, CO, 1997; [TU-97-3].
5. PECI. *What Can Commissioning Do for Your Building?* Portland Energy Conservation, Inc. for the U.S. Department of Energy: Washington, DC, 1997.
6. Texas A&M. *Continuous Commissioning Guidebook for Federal Energy Managers*, 2002. FEMP. Available URL: http://www.eere.energy.gov/femp/operations_maintenance.
7. Haasl, T., Sharp, T. A Practical Guide for Commissioning Existing Buildings, 1999. ORNL/TM-1999/34, Oak Ridge, TN.
8. Sydlowski, R.F. *Advanced Metering Techniques*; Pacific Northwest National Laboratory: Richland, WA, 1993; [PNL-8487].
9. AEC. *Advanced Utility Metering*; Architectural Energy Corporation: Boulder, CO, 2002; [Under contract NREL/SR-710-33539].
10. EPRI. *End-Use Performance Monitoring Handbook*; Electric Power Research Institute: Palo Alto, CA, 1996; [EPRI TR-106960].

Fossil Fuel Combustion: Air Pollution and Global Warming☆

Dan Golomb
Department of Environmental, Earth and Atmospheric Sciences, University of Massachusetts—Lowell, Lowell, Massachusetts, U.S.A.

Abstract
Currently, fossil fuels supply about 86% of the global primary energy consumption for transportation, industrial, commercial and residential uses. Due to the combustion of fossil fuels, copious quantities of pollutants are emitted into the air, which impact the local, regional and global air quality. Regulatory agencies in most developed countries prescribe ambient standards (concentrations) for air pollutants at a level above which the pollutants could have an adverse impact on human health and the ecology. These agencies also prescribe emission standards for various industrial categories and for automobiles at a level that is considered safe for human health and the environment. The relationship between emissions and ambient concentrations is formulated by atmospheric models that take into account emission rates from various sources, transport by winds, dispersion by turbulence, and the removal processes of the pollutants by wet and dry deposition. It is technologically possible to significantly reduce the emissions of most pollutants by employing various pollution abatement technologies. The reduction of emissions of carbon dioxide, the major contributor to global warming, poses technological, economic, societal and political challenges of enormous magnitude.

INTRODUCTION

Currently, fossil fuels supply about 86% of global primary energy consumption (39% oil, 24% coal, and 23% natural gas), providing energy for transportation, electricity generation, and industrial, commercial and residential uses. The air emissions of fossil fuel combustion are transported by winds and dispersed by atmospheric turbulence, eventually falling or migrating to the surface of the earth or ocean at various rates. While in the atmosphere, pollutants cause considerable harmful effects on human health, animals, vegetation, structures, and aesthetics.

In recent decades, it has become evident that rising levels of atmospheric carbon dioxide (and other greenhouse gases) have already warmed the earth's surface slightly. It is predicted that, with continuous and increasing use of fossil fuels, the warming trend will increase. Most of the increase in carbon dioxide levels is a direct consequence of fossil fuel combustion.

The goal of this article is to describe the characteristics of fossil fuel-generated air pollutants, including carbon dioxide and other greenhouse gases, their transport and fate in the atmosphere, and their effects on human health, the environment, and global climate change.

AMBIENT AND EMISSION STANDARDS OF AIR POLLUTANTS

Air pollutants, when they exceed certain concentrations, can cause acute or chronic diseases in humans, animals, and plants. They can impair visibility, cause climatic changes, and damage materials and structures. Regulatory agencies in most developed countries prescribe ambient standards (concentrations) for air pollutants. The standards are set at a level below which it is estimated that no harmful effect will ensue to human health and the environment. On the other hand, if the standards are exceeded, increased human mortality and morbidity, as well as environmental degradation, is expected. For example, in the United States, the Environmental Protection Agency (EPA) sets National Ambient Air Quality Standards (NAAQS) for six pollutants, the so-called criteria pollutants: carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), particulate matter (PM), and lead (Pb). The standards are listed in Table 1. In the past, particulate matter was regulated regardless of size (the so-called Total Suspended Particles, TSP). Starting in 1978, particles were regulated with an aerodynamic diameter less than 10 μm. This is called PM_{10}. In 1997, a new standard was introduced for particles with an aerodynamic diameter less than 2.5 μm. This is called $PM_{2.5}$. The reason for regulating only small particles is that these particles can be

☆ Text, figures, and tables for this article are excerpted from Golomb D.S.; Fay, J.A., *Atmospheric Impact of the Fossil Fuel Cycle*, Energy, Waste and the Environment: a Geochemical Perspective, Giere R., P. Stille, Eds.; Geological Society, Special Publications: London, 236, 2004. with permission from the Geological Society.

Keywords: Air pollution; Dry deposition; Fossil fuels; Global warming; Pollutant transport and dispersion; Pollution abatement; Wet deposition.

Table 1 U.S. 2000 National ambient air quality standards (NAAQS)

Pollutant	Standard	Averaging time
Carbon monoxide	9 ppm (10 mg/m^3)	8-h[a]
	35 ppm (40 mg/m^3)	1-h[a]
Lead	1.5 $\mu g/m^3$	Quarterly Average
Nitrogen dioxide	0.053 ppm (100 $\mu g/m^3$)	Annual (arithmetic mean)
Particulate matter (PM_{10})	50 $\mu g/m^3$	Annual[b] (arithmetic mean)
	150 $\mu g/m^3$	24-h[a]
Particulate matter ($PM_{2.5}$)	15.0 $\mu g/m^3$	Annual[c] (arithmetic mean)
	65 $\mu g/m^3$	24-h[d]
Ozone	0.08 ppm	8-h[e]
Sulfur oxides	0.03 ppm	Annual (arithmetic mean)
	0.14 ppm	24-h[a]
	—	3-h[a]

[a] Not to be exceeded more than once per year.
[b] To attain this standard, the 3-year average of the weighted annual mean PM_{10} concentration at each monitor within an area must not exceed 50 $\mu g/m^3$.
[c] To attain this standard, the 3-year average of the weighted annual mean $PM_{2.5}$ concentrations from single or multiple community-oriented monitors must not exceed 15.0 $\mu g/m^3$.
[d] To attain this standard, the 3-year average of the 98th percentile of 24-h concentrations at each population-oriented monitor within an area must not exceed 65 $\mu g/m^3$.
[e] To attain this standard, the 3-year average of the fourth-highest daily maximum 8-h average ozone concentrations measured at each monitor within an area over each year must not exceed 0.08 ppm.

lodged deep in the alveoli of the lungs, and hence are detrimental to our health, whereas the larger particles are filtered out in the upper respiratory tract. While there is a standard for lead concentrations in the air, this pollutant is no longer routinely monitored. With the phasing out of leaded gasoline in the 1970s, lead concentrations in the air steadily declined, and allegedly, in the United States, airborne lead no longer poses a health hazard. (Of course, lead in its other forms, such as in paint, pipes, groundwater, solder, and ores, is still a hazard.)

In addition, the EPA sets emission standards for stationary and mobile sources. Stationary sources include power plants, incinerators, steel, cement, paper and pulp factories, chemical manufacturers, refineries, and others. Mobile sources include automobiles, trucks, locomotives, ships, and aircraft. As an example, Table 2 lists the U.S. emission standards (New Source Performance Standards) for fossil-fueled steam generators, which include large fossil-fueled power plants and industrial boilers that were constructed after 1970. The emission standards are set in terms of mass emitted per fuel heat input (g/GJ). However, in 1978, new regulations were implemented. Instead of numerical emission standards, EPA prescribed *emission control technologies*, the so-called best available control technologies (BACT), purporting to reduce emissions to a minimum with practical and economic pollution abatement devices. For example, current BACT for sulfur oxides is a "scrubber," usually employing a slurry of pulverized limestone ($CaCO_3$) counter-flowing to the flue gas that contains the sulfur oxides. For control of nitric oxides (NO_x), the current BACT is the low-NO_x-burner (LNB). Because the LNB reduces NO_x emissions only up to 50%, new sources may be required to employ more efficient NO_x-reducing technologies, such as selective catalytic reduction (SCR) or selective non-catalytic reduction (SNCR), based on ammonia or urea injection into the flue gas. For control of particulate matter, the current prescribed technology is the electrostatic precipitator (ESP). Because the ESP is not efficient in removing sub-micron and 1–2 μm particles, and in view

Table 2 U.S. new source performance standards (NSPS) for fossil fuel steam generators with heat input > 73 MW

Pollutant	Fuel	Heat input (g/GJ)
SO_2	Coal	516
	Oil	86
	Gas	86
NO_2	Coal (bituminous)	260
	Coal (subbituminous)	210
	Oil	130
	Gas	86
PM[a]	All	13

[a] PM, particulate matter. For PM emissions an opacity standard also applies, which allows a maximum obscuration of the background sky by 20% for a 6-min period.
Source: Data from EPA. Standards of performance for new stationary sources. Electric steam generating units. *Federal Register*, 45, February 1980, 8210–8213.

Table 3 U.S. federal vehicle emission standards

	Light-duty vehicles (auto)				Light-duty trucks (gasoline)			
Model year	HC (g/km)	CO (g/km)	NO_x (g/km)	PM (g/km)	HC (g/km)	CO (g/km)	NO_x (g/km)	PM (g/km)
1968	2.00	20.50						
1971[a]	2.90	29.20	2.49					
1974	2.11	24.20	1.86					
1977	0.93	9.32	1.24					
1978	0.93	9.32	1.24		1.24	12.40	1.93	
1979	0.93	9.32	1.24		1.06	11.20	1.43	
1980	0.25	4.35	1.24		1.06	11.20	1.43	
1981	0.25	2.11	0.62		1.06	11.20	1.43	
1982	0.25	2.11	0.62	0.37	1.06	11.20	1.43	
1985	0.25	2.11	0.62	0.37	0.50	6.21	1.43	0.99
1987	0.25	2.11	0.62	0.12	0.50	6.21	1.43	1.62
1988	0.25	2.11	0.62	0.37	0.50	6.21	0.75	1.62
1994	0.25	2.11	0.25	0.05	0.50	3.42	0.60	0.06

[a] Test method changed in 1971.

of the new $PM_{2.5}$ standard, new sources may be required to install instead of an ESP a Fabric Filter (FF), also called a bag house. The FF consists of a porous fabric or fiberglass membrane that filters out efficiently the smallest particles, albeit at an increased cost and energy penalty compared to the ESP. The detailed description of the workings of these emission control technologies is beyond the scope of this article; the reader is referred to the excellent handbooks on the subject.[1,2]

Table 3 lists the U.S. emission standards in units of mass emitted per length traveled (g/km) for passenger cars and light trucks for the different model years from 1968 to 1994. At present, the 1994 standards are still in effect. Notice that light truck standards are not as strict as those for light duty vehicles. This dichotomy is very controversial, because the light truck category includes sport utility vehicles (SUV), minivans, and pick-up trucks, which are mostly used for personal and not for cargo transport. Therefore, their emission standards ought to be equal to those for passenger cars. Until now, these vehicles captured the majority of sales in the United States. (With rising gasoline prices, this trend may be reversed.) The achievement of the emission standards relies on emission control technologies. For unleaded gasoline-fueled vehicles, the prevailing control technology is the three-way catalytic converter, which simultaneously reduces carbon monoxide (CO), nitric oxides (NO_x), and fragmentary hydrocarbon (HC) emissions. Unfortunately, the catalytic converter would not work on diesel-fueled vehicles, because the relatively high sulfur and particle emissions from diesel engines would poison the catalyst. Automobile and truck manufacturers are intensively investigating possible technologies that would reduce emissions from diesel-fueled autos and trucks.

In addition to the aforementioned criteria pollutants, one finds in the air a host of other gaseous and particulate pollutants, generally designated as hazardous air pollutants (HAP), or simply *air toxics*. The EPA has identified 189 HAPs. Of course, not all HAPs are related to fossil fuel usage. Examples of fossil fuel HAPs are products of incomplete combustion (PIC), volatile organic compounds (VOC), polycyclic aromatic hydrocarbons (PAH), toxic metals (e.g., mercury, cadmium, selenium, arsenic, vanadium, etc.) Many of the fossil fuel related HAPs are found as condensed matter on particles emitted by stationary sources (e.g., fly ash from power plants) or mobile sources (e.g., exhaust smoke from trucks). While HAPs may be more harmful to our health than the criteria pollutants (some of them are carcinogens), it is difficult to establish a dose–response relationship. Therefore, instead of setting HAP emission standards, the EPA mandates that specific control technologies be installed on major emitting sources. These are called maximum achievable control technologies (MACT). For example, for toxic volatile gases, the most often used control technologies are physical adsorption on porous adsorbents (usually activated carbon), or chemical absorption in solvents, including water, that have a large absorption capacity for these gases. When the concentration of toxic gases in the effluent is high, secondary incineration of the effluent may be warranted. For detailed descriptions of MACTs, the reader is also referred to the appropriate handbooks.[1,2]

POLLUTANT TRANSPORT AND DISPERSION

When air pollutants exit a smoke stack or exhaust pipe (called the *sources*), they are transported by winds and dispersed by turbulent diffusion. Winds blow from high pressure toward low-pressure cells at speeds that depend on the pressure gradient. Because of the Coriolis force, wind trajectories are curvilinear in reference to fixed earth coordinates, although within a relatively short (few to tens of km) distance, wind trajectories can be approximated as linear. Winds have a horizontal and vertical component. Over flat terrain, the horizontal component predominates; in mountainous and urban areas with tall buildings, the vertical component can be significant, as well at the land/sea interface.

In the bottom layer of the atmosphere, called the troposphere, the temperature usually declines with altitude. For a dry atmosphere, the temperature gradient is $-9.6°C/km$. This is called the *dry adiabatic lapse rate*. In a moist atmosphere, the temperature gradient is less. Under certain conditions (e.g., nocturnal radiative cooling of the earth surface), the temperature gradient may even become positive. This is called *inversion*. The global and temporal average temperature gradient in the troposphere is $-6.5°C/km$. When a negative temperature gradient exists, upper parcels of the air are denser than lower ones, so they tend to descend, while lower parcels buoy upward, giving rise to eddy or turbulent mixing. The more negative the temperature gradient, the stronger the turbulence, and the faster the dispersion of a pollutant introduced into the troposphere. Wind shears also cause turbulence. A wind shear exists when, in adjacent layers of the atmosphere, winds blow in different directions and speeds.

During an inversion, there is little turbulence, and a pollutant will disperse very slowly. Inversions may also occur aloft, that is, a negative temperature gradient exists at the bottom, followed by a positive gradient above. The layer up to the altitude at which the inversion occurs is called the *mixing layer* and the altitude at the inversion is called the *mixing height*. The shallower the mixing layer, the greater chances for air pollution episodes to develop, because pollutants emitted near the ground are confined to the shallow mixing layer. This often occurs in cities such as Los Angeles, Houston, Atlanta, Salt Lake City, Denver, Mexico City, Sao Paulo, Athens, Madrid, Rome, and Istanbul. Some of these cities are surrounded by mountain chains. In the basin of the mountain chain, or in the valley, the mixing layer is shallow, and the winds in the layer are usually weak, leading to poor ventilation. During the morning rush hour, pollutants are emitted into the shallow mixing layer, where they are concentrated because of the "lid" imposed by the inversion aloft. Later in the day, when solar radiation breaks up the inversion, the pollutants disperse to higher altitude, and the pollutant concentration becomes more diluted. The photo in Fig. 1 shows a pollution episode in Los Angeles. The mixing height extends only to the middle of the tall building. The lower part is obscured by the smog, the upper part, which is above the mixing layer, is in relatively clear air.

Fig. 1 Los Angeles smog. Note the low inversion beneath which the smog accumulates. Above the inversion, the air is relatively clear. Photo by the South Coast Air Quality Management District.

Air Quality Modeling

The estimation of ambient pollutant concentrations in space and time due to emissions of single or multiple sources is called air quality modeling (AQM), or source–receptor modeling (SRM). The basic ingredients of AQM are the emission strengths of the sources, meteorological conditions, and solar irradiation. Air quality models are of the trajectory-type, where the coordinate system moves together with the plume of pollutants, or grid-type, where the coordinate system is fixed over an area, and the emission strengths and meteorological variables are inserted in each grid of the model domain. Some models consider only conservative pollutants, where the emitted pollutant does not change en route to the receptor. Other models consider chemical transformation processes and "sinks," e.g., dry and wet deposition to the ground. Air quality models where transformation and deposition processes need to be considered are acid deposition, regional haze, and photo-oxidants.

Acid Deposition

While commonly called acid rain, acid deposition is a better term, because deposition can occur both in the wet and dry form. The ingredients of acid deposition are sulfuric and nitric acids. The primary pollutants are SO_2 and NO_x, which is the sum of NO and NO_2 molecules. Both SO_2 and NO_x result from fossil fuel combustion. Sulfur is a ubiquitous ingredient of coal and petroleum. When these fuels are burned in air, SO_2 is emitted from a smoke stack or the exhaust pipe of a vehicle. Coal and petroleum also contain nitrogen in their molecular make-up, resulting in NO_x emissions when these fuels are combusted in air. In addition, some NO_x is formed from the recombination of air oxygen and nitrogen at the high flame temperatures. So, even the combustion of natural gas, which has no nitrogen in its molecular make-up, produces some NO_x.

Primary emitted SO_2 and NO_x are transformed in the atmosphere to sulfuric and nitric acids. The resulting acids can be either deposited in the dry phase on land or water, a process called dry deposition, or scavenged by falling rain drops or snow flakes, resulting in wet deposition.[3] Acid deposition modeling is successful in predicting the amount of acid deposition given the emission strength of the precursors SO_2 and NO_x.[4] From these models, it was concluded that acid deposition within a geographic domain is approximately linearly dependent on SO_2 and NO_x emission strength in that domain, so a certain percentage of precursor emission reduction results in a proportional deposition reduction of sulfuric or nitric acid. Indeed, in countries and continents where serious curtailments of precursor emissions have been made, a proportional reduction of acid deposition occurred. In the United States, as a result of reducing emissions of SO_2 by approximately one-half since the enactment of the Clean Air Act of 1990, sulfuric acid deposition has declined by approximately one-half.[5] Nitric acid deposition has not fallen appreciably, because the control of NO_x is much more difficult to accomplish than the control of SO_2, especially from dispersed sources such as commercial and residential boilers and furnaces, automobiles, and diesel trucks.

Regional Haze

Small particles (also called fine particles or aerosols), less than 1–2 μm in diameter, settle very slowly on the ground. Small particles can be either in the solid or liquid phase. For example, fly ash or smoke particles are solid; mist is liquid. Small particles can be of natural origin (e.g., volcanic dust, wind blown soil dust, forest and brush fires) or of anthropogenic origin (e.g., fly ash, diesel truck smoke). They can be emitted as primary particles or formed by transformation and condensation from primary emitted gases. They can travel hundreds to thousands of kilometers from their emitting sources. The particles can envelope vast areas, such as the northeastern or southwestern United States, southeastern Canada, western and central Europe, and southeastern Asia. Satellite photos often show continental areas covered with a blanket of haze. The haze may extend far out over the ocean. This phenomenon is called regional haze. Small particles are efficient scatterers of sunlight. Light scattering prevents distant objects from being seen. This is called visibility impairment. Fig. 2 shows an encroaching haze episode on a mountain chain in Vermont. As the haze thickens, distant mountain peaks are no longer visible, and eventually neighboring peaks disappear. Increasing concentration of particles in urbanized parts of continents causes the loss of visibility of the starlit nocturnal sky. These days, small stars, less than the fifth order of magnitude, rarely can be seen from populated areas of the world.

The composition of fine particles varies from region to region, depending on the precursor emissions. In the northeastern United States, central Europe, and southeastern Asia, more than half of the composition is made up of sulfate particles, due to the combustion of high sulfur coal and oil. The rest is made up of nitrate particles, carbonaceous material (elemental and organic carbon), and crustal matter (fugitive particles from soil, clay, and rock erosion).

Photo-Oxidants

The family of photo-oxidants includes tropospheric ozone, O_3 (the bad ozone), ketones, aldehydes, and nitrated oxidants, such as peroxyacetylnitrate (PAN) and peroxybenzoylnitrate (PBN). The modeling of photo-oxidants is more complicated than that of acid deposition.[6] Here, the primary precursor is NO_x, which as mentioned previously,

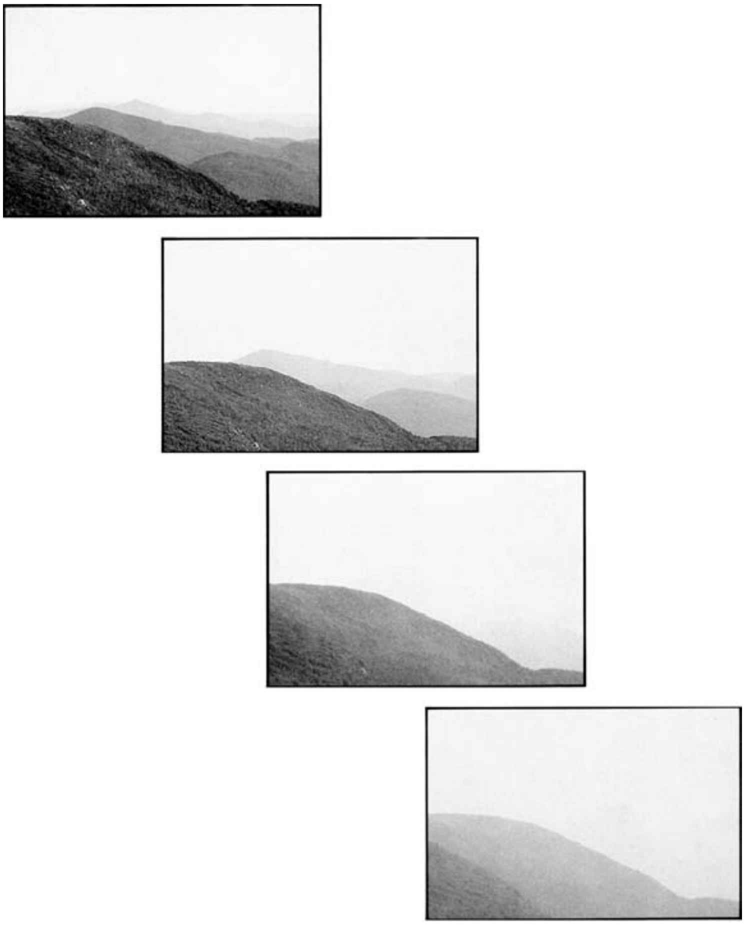


Fig. 2 Regional haze in the Vermont green mountains. Note that as the pollution episode progresses, adjacent peaks are no longer visible. Photo by the Vermont agency of environmental conservation.

is emitted because of fossil fuel combustion. A part of NO_x is the NO_2 molecule, which splits (photo-dissociates) by solar ultraviolet and blue photons into NO and atomic oxygen. The photo-dissociation rate is dependent on solar irradiation, which, in turn, is dependent on latitude, season, time of day, and cloudiness. Atomic oxygen combines with molecular oxygen to form O_3. The NO that is formed in the photo-dissociation is quickly re-oxidized into NO_2 by peroxy radicals, RO_2, present in the polluted atmosphere. The peroxy radicals are formed from VOCs that are emitted as a consequence of incomplete combustion of fossil fuels, or from evaporation and leakage of liquid fossil fuels and solvents. The VOCs are oxidized in a complicated sequence of photo-chemical reactions to the peroxy radicals. The oxidation rates of VOCs are also dependent on solar irradiation and on the specific VOC molecule. Long- and branch-chained hydrocarbons (e.g., *n*-octane and *iso*-octane) are more reactive then short- and straight-chained ones (e.g., methane and ethane); unsaturated hydrocarbons (e.g., ethene) are more reactive than saturated ones (e.g., ethane). Aromatic hydrocarbons (e.g., benzene) are more reactive than aliphatic ones (e.g., hexane), and so on.

Thus, photo-oxidants have two kinds of precursors, NO_x and VOCs, which make abatement of these secondary pollutants, as well as their modeling, so complicated. First of all, as mentioned previously, complete NO_x emission control is difficult to accomplish, because in addition to coming from large stationary sources, NO_x is emitted from a myriad of dispersed sources, such as home and commercial furnaces and boilers, automobiles, trucks, off-road vehicles, aircraft, locomotives, and ships. In principle, anthropogenic VOCs could be substantially controlled, for example, by ensuring complete combustion of the fossil fuel, or with the catalytic converter on automobiles. But, not all VOCs are of anthropogenic origin. Trees and vegetation emit copious quantities of VOCs, such as terpenes and pinenes, which are pleasant smelling, but they do participate in photo-chemical reactions that produce photo-oxidants. Even though great effort and expenses are being made in many developed countries to control precursors, photo-oxidant

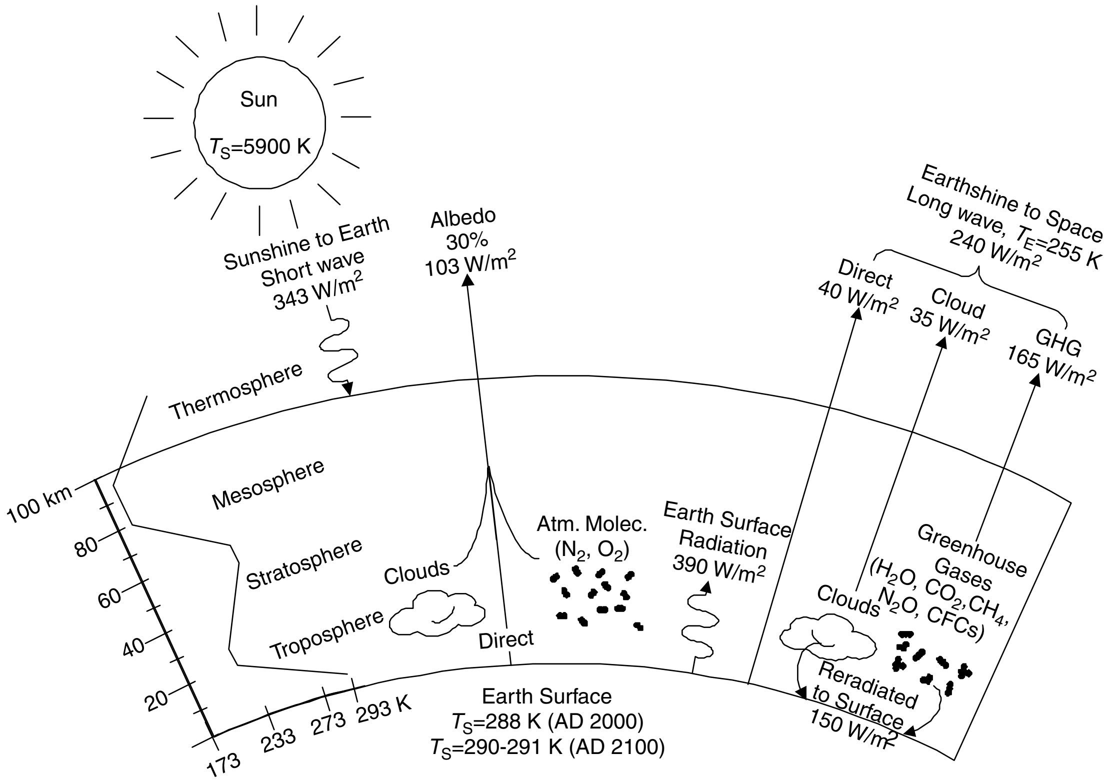


Fig. 3 Schematic of the greenhouse effect.

concentrations over urban-industrial continents have improved only slightly, if at all. In less developed countries that do not have the means of controlling photo-oxidant precursors, their concentrations are on a steady increase.

GLOBAL WARMING

Of all environmental effects of fossil fuel usage, global warming, including its concomitant climate change, is the most perplexing, potentially most threatening, and arguably most intractable problem. It is caused by the ever-increasing accumulation in the atmosphere of carbon dioxide (CO_2) and other gases, such as methane (CH_4), nitrous oxide (N_2O), and chloro-fluoro-carbons (CFC), collectively called greenhouse gases (GHG). Atmospheric aerosols, natural as well as anthropogenic, also may contribute to global warming.

The term greenhouse effect is derived by analogy to a garden greenhouse. There, a glass-covered structure lets in the sun's radiation, warming the soil and plants that grow in it, while the glass cover restricts the escape of heat into the ambient surroundings by convection and radiation. Similarly, the earth atmosphere lets through most of the sun's radiation, which warms the earth surface, but the GHGs and some aerosols trap outgoing terrestrial infrared (IR) radiation, keeping the earth's surface warmer than if the GHGs and aerosols were absent from the atmosphere.

A schematic of the greenhouse effect is represented in Fig. 3. To the left of the schematic is the atmospheric temperature structure, which defines the various spheres. In the troposphere, the temperature decreases on average by 6.5°C/km. This sphere is thoroughly mixed by thermal and mechanical turbulence, and this sphere contains most of the air pollutants mentioned in the previous sections. In the stratosphere, the temperature increases steadily. Because of the positive temperature gradient, this sphere is very stable with very little turbulence and mixing. At about 50–60 km height, the temperature declines again with altitude, giving rise to the mesosphere. Finally, above 90–100 km, the temperature gradient reverses itself, and becomes positive. The highest sphere is called the thermosphere. Its temperature can reach hundreds to one thousand degrees, depending on solar radiation intensity that directly heats this sphere.

The average solar radiation that impinges on the top of the atmosphere is about 343 W/m^2. This is the annual, diurnal, and spatial average irradiation. Of this irradiation, currently about 30% (103 W/m^2) is immediately reflected into space by land, ocean, icecaps, and clouds, and scattered into space by atmospheric molecules. The reflected and scattered sunlight is called albedo. The albedo may not remain constant over time. With increased melting of the ice caps, and increased cloud cover, in part due to anthropogenic influences, the albedo may change over time. The remaining 70% of solar irradiation heats the earth's surface, land, and oceans. Currently, the global average surface temperature is 288 K (about 15°C). A body (the so-called black body) that is heated to 288 K radiates 390 W/m^2. The earth's surface radiation occurs in the far IR. This is called earth shine. A part of the earth shine is reflected back to the earth's surface by clouds and aerosols; another part is first absorbed by certain gaseous molecules, and then re-radiated back to the surface. The absorption/re-radiation occurs by poly-atomic molecules, including water vapor and the GHGs: CO_2, CH_4, N_2O, O_3, CFCs, and others. The reflection and re-radiation to the earth's surface of the outgoing terrestrial IR radiation is causing the earth's surface to become warmer than it would be merely by solar irradiation. This is the greenhouse effect. With increasing concentrations of anthropogenic GHGs and aerosols, the earth's surface temperature will increase. It should be emphasized that the greenhouse effect is not due to trapping of the incoming solar radiation, but outgoing terrestrial IR radiation. In fact, because of the trapping of earth shine by GHGs, the upper layers of the atmosphere (primarily the stratosphere) will become colder, not warmer.[7]

The extent of global warming can be predicted by radiative transfer models. These models include the radiative properties of GHGs and their distribution in the earth's atmosphere, as well as the temperature and pressure gradients in the atmosphere. There is general agreement among the models as to the extent of surface warming due to GHG absorption/re-radiation, called radiative forcing. Based on the models, it is predicted that the average earth surface temperature will increase as shown in Fig. 4. The middle, "best," estimate predicts a rise of the earth's surface temperature by the end of the 21st century of about 2°C; the "optimistic" estimate predicts about 1°C, and the "pessimistic" estimate predicts about a 3°C rise. The optimistic estimate relies on the slowing of CO_2 and other GHG emissions; the pessimistic estimate relies on "business-as-usual," i.e., on the continuing rate of growth of CO_2 and other GHG emissions, and the "best" estimate is somewhere in between.[8] If the GHG concentrations increase in the atmosphere at their current rate, by the year 2100, CO_2 will contribute about 65% to global warming, CH_4 15%, N_2O 10%, and CFC about 5%–10%. (By international conventions, CFCs are being phased out entirely. But because of their very long lifetime in the atmosphere, they still will contribute to global warming by the year 2100.)

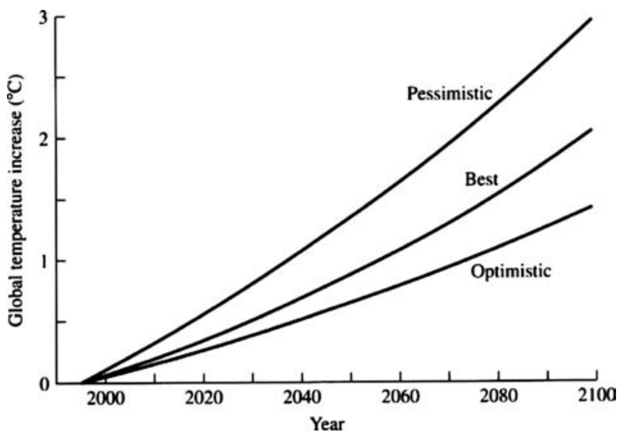


Fig. 4 Projected trend of the earth surface temperature increase. Upper curve: pessimistic scenario with no emission curtailment; lower curve: optimistic scenario with significant emission curtailment; middle curve: in-between scenario.
Source: From Cambridge University Press (see Ref. 8).

In addition to radiative forcing, global warming may be enhanced by the so-called feedback effects. For example, water vapor is a natural GHG. When the temperature of the ocean surface increases, the evaporation rate will increase. As a consequence, the average water vapor content of the atmosphere will increase. This causes more absorption of the outgoing infrared radiation and more global warming. Furthermore, increased evaporation may cause more cloud formation. Clouds and aerosols also can trap outgoing terrestrial radiation, further increasing global warming. Melting icecaps and glaciers decrease the reflection of incoming solar radiation (reduced albedo), which also increases global warming. The prediction of the feedback effects is more uncertain than the prediction of radiative forcing, but generally, it is assumed that the feedback effects may double the surface temperature increases because of radiative forcing alone.

Has the surface temperature already increased due to anthropogenic activities? Fig. 5 plots the global average surface temperature over the last century and a half. Even though there are large annual fluctuations, the smoothed curve through the data points indicates an upward trend of the temperature. From 1850, the start of the Industrial Revolution, to date, the global average surface temperature increased by about 0.5°C–1°C. This is in accordance with radiative models that predict such a trend, considering the increase of GHG concentrations over that period.

Other Effects of Global Warming

Because of increased GHG concentrations in the atmosphere, the earth's surface temperature may rise, as discussed in the previous section. The surface temperature rise may cause several ancillary effects on global climate and hydrogeology, which in turn will affect human habitat, welfare, and ecology.

Sea Level Rise

With increasing surface temperatures, the average sea level will rise because of three factors: melting of polar ice caps, receding of glaciers, and thermal expansion of the

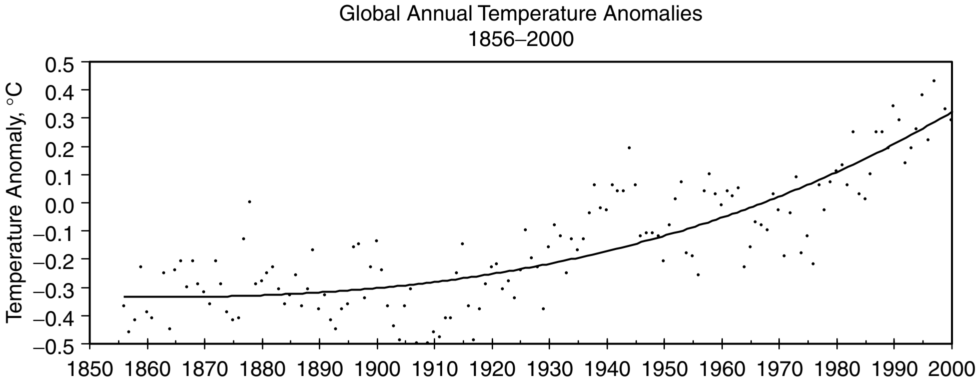


Fig. 5 Global average surface temperature trend 1850–2000. Source: From Oak Ridge National Laboratory (see Ref. 9).

ocean surface waters. Combining all three factors, it is estimated that by the end of the next century, the average sea level may be 30–50 cm higher than it is today. This can seriously affect low-lying coastal areas, such as The Netherlands in Europe, Bangladesh in Asia, and low-lying islands in the Pacific and other oceans.[8]

Climate Changes

Predicting global and regional climatic changes because of average surface temperature rise is extremely difficult and fraught with uncertainties. It is expected that regional temperatures, prevailing winds, and storm and precipitation patterns will change, but where and when, and to what extent changes will occur, is a subject of intensive investigation and modeling on the largest available computers, the so-called supercomputers. Climate is not only influenced by surface temperature changes, but also by biological and hydrological processes, and by the response of ocean circulation, which are all coupled to temperature changes. It is expected that temperate climates will extend to higher latitudes, probably enabling the cultivation of grain crops further toward the north than at present. But crops need water. On the average, the global evaporation and precipitation balance will not change much, although at any instant, more water vapor (humidity) may be locked up in the atmosphere. However, precipitation patterns may alter, and the amount of rainfall in any episode may be larger than it is now. Consequently, the runoff (and soil erosion) may be enhanced, and areas of flooded watersheds may increase. Hurricanes and typhoons spawn in waters that are warmer than 27°C, in a band from 5 to 20° north and south latitude. As the surface waters become warmer, and the latitude band expands, it is very likely that the frequency and intensity of tropical storms will increase.

The sea level and climatic changes may cause a redistribution of agricultural and forestry resources, a considerable shift in population centers, and incalculable investments in habitat and property protection.

Greenhouse Gas Concentrations Trends

Currently, about 6.8 Gt/y of carbon (25 Gt/y CO_2) are emitted into the atmosphere by fossil fuel combustion. Another 1.5 ± 1 Gt/y are emitted due to deforestation and land use changes, mainly artificial burning of rain forests in the tropics, and logging of mature trees, which disrupts photosynthesis. Fig. 6 plots the trend of atmospheric concentrations of CO_2, measured consistently at Mauna Loa, Hawaii since 1958 to date. At present, the average CO_2 concentration is about 375 parts per million by volume (ppmv). The plot shows seasonal variations due mainly to assimilation/respiration of CO_2 by plants, but there is a steady increase of the average concentration at a rate of approximately 0.4%/y. If that rate were to continue into the future, a doubling of the current CO_2 concentration would occur in about 175 years. However, if no measures are taken to reduce CO_2 emissions, then due to the population increase, and the concomitant enhancement of fossil fuel use, the rate of growth of CO_2 concentration will increase more than 0.4%/y, and the doubling time will be achieved sooner.

Methane emissions are in part due to fossil energy usage, because CH_4 leaks from gas pipes, storage tanks, tankers, and coal mine shafts. Anthropogenic emissions of CH_4 from fossil fuel usage amount to about 100 Mt/y. However, CH_4 is also emitted from municipal waste landfills, sewage treatment, biomass burning, cultivated rice paddies, enteric fermentation of cattle, and other anthropogenic activities, so that the total amount of CH_4 emissions is about 400 Mt/y. Currently, the average atmospheric concentration of CH_4 is about 1.7 ppmv, growing at about 0.6%/y. Nitrous oxide (N_2O) is a minor product of combustion of fossil fuels. Currently, the concentration of N_2O is about 0.3 ppmv, growing by about 0.25%/y. CFCs are not directly associated with fossil fuel

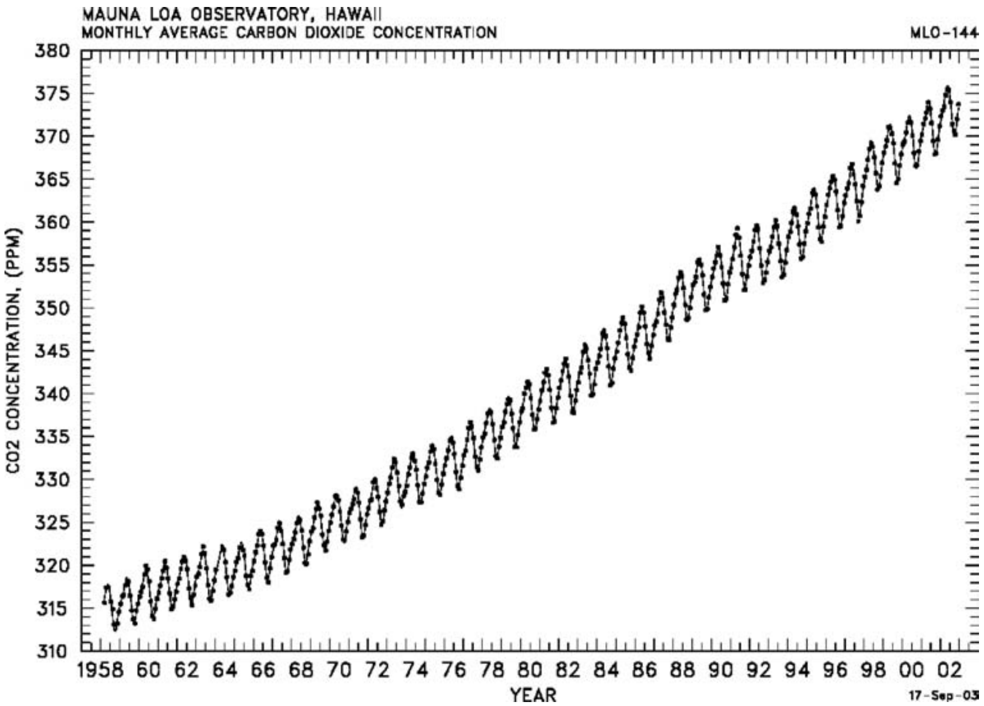


Fig. 6 Carbon dioxide concentration trend 1958–2002.
Source: From Oak Ridge National Laboratory (see Ref. 9).

usage; however, they are emitted inadvertently from energy using devices, such as refrigerators, air conditioners, chillers, and heat pumps. Current concentrations of the various CFCs are about 0.5 parts per billion by volume (ppbv), and their concentrations are slowly declining due to the phase-out of production of CFCs.

What Can be Done about Global Warming

Global warming could be lessened by reducing significantly the emissions of GHGs into the atmosphere. Most GHG emissions are a consequence of fossil fuel use. While it is important to reduce the emissions of all GHGs, the greatest preventative measure would come from reducing CO_2 emissions. CO_2 emission reductions can be accomplished by a combination of several of the following approaches.

Demand-side Conservation and Efficiency Improvements

This includes less space heating and better insulation, less air conditioning, use of more fluorescent instead of incandescent lighting, more energy efficient appliances, process modification in industry, and very importantly, more fuel efficient automobiles. Some measures may even incur a negative cost, i.e., consumer savings by using less energy, or at least a rapid payback period for the investment in energy saving devices.

Supply-side Efficiency Measures

Here we mean primarily increasing the efficiency of electricity production. Natural gas combined cycle power plants (NGCC) emit less CO_2 than single cycle coal fired power plants. First, because NG emits about one-half the amount of CO_2 per fuel heating value than coal, and second, because the thermal efficiency of combined cycle power plants is in the 45%–50% range vs the 35%–38% range of single cycle plants. In the future, integrated coal gasification combined cycle power plants (IGCC) may come on-line with a thermal efficiency in the 40%–45% range reckoned on the basis of the coal heating value. Furthermore, IGCC may enable the capture of CO_2 at the gasification stage, with subsequent sequestration of the captured CO_2 in geologic and deep ocean repositories.

Shift to Non-Fossil Energy Sources

The choices here are agonizing, because the largest impact could be made by shifting to nuclear and hydro electricity, both presently very unpopular and fraught with environmental and health hazards. The shift to solar, wind, geothermal, and ocean energy is popular, but because of their limited availability and intermittency, and because of their larger cost compared to fossil energy, a substantial shift to these energy sources cannot be expected in the near future.

None of these options can prevent global warming by itself. They have to be taken in combination and on an incremental basis, starting with the least expensive ones and progressing to the more expensive ones. Even if the predictions of global climate change were to turn out exaggerated, the fact that fossil fuel usage entails many other environmental and health effects, and the certainty that fossil fuel resources are finite, makes it imperative that we curtail fossil energy usage as much as possible.

SUMMARY AND CONCLUSIONS

The use of fossil fuels to supply energy for the use of the world's population has resulted in the release to the atmosphere of troublesome chemical byproducts that present harm to humans and other natural species. These effects can be localized (near the emission source), can extend to large regional areas (involving sub-continents), and can even cover the globe, from pole to pole. A large portion of the human population is exposed to one or more of these environmental effects. The scientific understanding of how fossil fuel use causes these effects is well advanced, providing quantitative means for explaining what is currently observed and predicting what changes will occur in the future from projected future fuel consumption. These projections provide a basis for modifying the amount and character of future energy supply so as to lessen harmful environmental consequences. The technological systems that employ fossil fuel energy have been developed to lessen the amounts of harmful emissions, albeit at significant energy and economic cost. Further improvements can be expected, but at increasing marginal cost. The most severe emission control problem, in terms of economic and energy cost, is CO_2, a major contributor to global warming. The implementation of policies by national governments and international bodies to curtail the use of fossil energy and the concomitant emissions of CO_2 will become a growing task for humankind in this century.

REFERENCES

1. Wark, K.; Warner, C.F.; Davis, W.T. *Air Pollution: Its Origin and Control*, 3rd Ed.; Addison-Wesley: Menlo Park, CA, 1998.
2. Cooper, C.D.; Alley, F.C. *Air Pollution Control: A Design Approach*, 3rd Ed.; Waveland Press, Prospects Heights: IL, 2002.
3. Seinfeld, J.H.; Pandis, S.N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; Wiley: New York, 1998.
4. Fay, J.A.; Golomb, D.; Kumar, S. Source apportionment of wet sulfate deposition in Eastern North America. Atm. Environ. **1985**, *19*, 1773–1782.
5. Butler, T.J.; Likens, G.E.; Stunder, B.J.B. Regional-scale impacts of phase I of the clean air act amendments in the U.S.A.: the relation between emissions and concentrations, both wet and dry. Atmos. Environ. **2000**, *35*, 1015–1028.
6. National Research Council *Rethinking the Ozone Problem in Urban and Regional Air Pollution*; National Academy Press: Washington, DC, 1991.
7. Mitchell, J.F.B. The greenhouse effect and climate change. Rev. Geophys. **1989**, *27*, 115–139.
8. Intergovernmental Panel on Climate Change *Climate Change: The Scientific Basis*; Cambridge University Press: Cambridge, U.K., 2001.
9. CO_2 Information Center *Trends Online*; Oak Ridge National Laboratory: Oak Ridge, TN, 2000.

Fuel Cells: Intermediate and High Temperature

Xianguo Li

Gholamreza Karimi

Department of Mechanical Engineering, University of Waterloo, Waterloo, Ontario, Canada

Abstract

This entry provides an overview of intermediate- and high-temperature fuel cells, including phosphoric acid fuel cells, molten carbonate fuel cells, and solid oxide fuel cells. For each of them, a brief description is given of their operational characteristics, acceptable contamination levels, major technical barriers to commercialization, technological status, economics, and major intended applications.

INTRODUCTION

The three major types of low-temperature fuel cells were described in the previous entry, titled "Low Temperature Fuel Cells." This entry will elaborate on intermediate- and high-temperature fuel cells, including phosphoric acid fuel cells (PAFCs), molten carbonate fuel cells (MCFCs), and solid oxide fuel cells (SOFCs). The PAFC is the most commercially developed fuel cell. It is being used for distributed or on-site cogeneration in such locations as hospitals, hotels, offices, and schools. It can also be used in large vehicles (buses, etc.). The operating temperature range of PAFC is between 160 and 220°C, and its efficiency is approximately 40% lower heating value (LHV). The high-temperature MCFCs and SOFCs boast advantages over conventional power-generating systems in terms of fuel flexibility, modularity, low emission of NO_x and SO_x pollutants, and environmental friendliness. They also demonstrate relatively high tolerance to trace levels of impurities in the gas stream. In addition, due to their high operating temperatures (typically in the range of 600°C–1000°C), hydrocarbon fuels such as methane and natural gas can be reformed within the stack, eliminating the need for expensive external reforming systems. The high operating temperatures require that most applications for these fuel cells be limited to large, stationary power plants, although SOFCs are also being considered as auxiliary powers for large transport vehicles (trucks). The high-quality heat produced in cogeneration power systems can be applied to space heating, industrial processing, and even steam turbines to generate more electricity, improving the system efficiencies to high levels (up to 70% [LHV]).

Keywords: Fuel cells; Phosphoric acid; Molten carbonate; Solid oxide.

PHOSPHORIC ACID FUEL CELLS

The phosphoric acid fuel cell (PAFC) is the most advanced type of fuel cell and is considered to be "technically mature" and ready for commercialization after nearly 40 years of research, development, and demonstration (RD&D) and an expenditure of over half a billion dollars. Therefore, the PAFC has been referred to as the first-generation fuel cell technology. Unlike the alkaline fuel cell systems, which were primarily developed for space applications, the PAFC was targeted initially for terrestrial applications with the carbon dioxide-containing air as the oxidant gas and hydrocarbon-reformed gas as the fuel for electrochemical reactions and electric power generation.

The basic components of a phosphoric acid fuel cell are the electrodes, consisting of finely dispersed platinum catalyst or carbon paper; an SiC matrix holding the phosphoric acid; and a bipolar graphite plate with flow channels for fuel and oxidant. The operating temperature ranges between 160 and 220°C, and it can use hydrogen, hydrogen-enriched gas produced from hydrocarbons (typically natural gas), or alcohols as the anodic reactant. In the case of hydrogen produced from a reformer with air as the anodic reactant, a temperature of 200°C and a pressure as high as 8 atm are required for better performance. Phosphoric acid fuel cells are advantageous from a thermal management point of view. The rejection of waste heat and product water is very efficient in this system, and the waste heat, at about 200°C, can be used efficiently for the endothermic steam-reforming reaction. The waste heat can also be used in cogeneration for space heating and hot water supply.

However, the PAFC cannot tolerate the presence of carbon monoxide and H_2S, which are commonly present in the reformed fuels. These contaminants poison the catalyst and decrease its electrochemical catalytic activity. A major challenge for using natural gas reformed fuel, therefore, lies in the removal of carbon monoxide to a level of less than 200–300 ppm. Carbon monoxide tolerance is higher at an operating temperature above 180°C. However, removal of sulfur is still essential. Further, the PAFC demonstrates an inferior performance, primarily due to the slow oxygen reduction reaction at the cathode. Therefore, the PAFC is typically operated at a higher temperature (near 200°C) for better electrochemical reactivity and smaller internal resistance, due mainly to the phosphoric acid electrolyte. As a result, PAFC exhibits the problems of both high- and low-temperature fuel cells, and perhaps none of the advantages of either option.

The PAFC system is the most advanced fuel cell system for terrestrial applications. Its major use is in on-site integrated energy systems to provide electrical power in apartments, shopping centers, office buildings, hotels, and hospitals. These fuel cells are commercially available in the range from 24 V/250 W portable units to 200 kW on-site generators. Phosphoric acid fuel cell systems of 0.5–1.0 MW are being developed for use in stationary power plants of 1–11 MW in capacity. The power density of a PAFC system is about 200 mW/cm^2, and the power density for a 36 kW brassboard PAFC fuel cell stack has been reported to be 0.12 kW/kg and 0.16 kW/L. The most advanced PAFC system is the PC-25 from International Fuel Cells in Connecticut (now it is named PureCell™ 200 under UTC Power). It costs about $3000 U.S./kW (the best technology possible for the PAFCs), while the conventional thermal power generation system costs only about $1000 U.S./kW. Thus, it is believed that the PAFC is not commercially viable, even though the U.S. Department of Energy (DOE) and the U.S. Department of Defense (DOD) have been subsidizing half of the cost to gain operational and maintenance experience for practical fuel cell systems. Although Japan seems determined to push ahead for this fuel cell technology, interest in the PAFC systems is waning in the United States and Europe.

MOLTEN CARBONATE FUEL CELLS

Introduction

The molten carbonate fuel cell (MCFC) is often referred to as the second-generation fuel cell because its commercialization is expected to come to fruition after that of the PAFC. It is believed that the development and technical maturity of the MCFC are approximately 5–7 years behind the PAFC. At present, the MCFC has reached the early demonstration stage of precommercial stacks, marking the transition from fundamental and applied research and development (R&D) to product development. Molten carbonate fuel cells are being targeted to operate on coal-derived fuel gases or natural gas. This contrasts with the previously discussed PAFCs, which prefer natural gas as primary fuel.

The MCFC operates at higher temperatures than all other fuel cells described so far. The operating temperature of the MCFC generally ranges from 600 to 700°C (typically 650°C). Such high temperatures produce high-grade waste heat, suitable for fuel processing, cogeneration, or combined cycle operation, and leading to higher electric efficiency. It also yields the possibility of utilizing carbonaceous fuels (especially natural gas) directly, through internal reforming to produce the fuel (hydrogen) ultimately used by the fuel cell electrochemical reactions. This results in simpler MCFC systems (i.e., without external reforming or fuel processing subsystems), less parasitic load, and less cooling power requirements, leading to higher overall system efficiency as well. The high operating temperature reduces voltage losses due to reduced activation, ohmic, and mass transfer polarization. The activation polarization is reduced to such an extent that it does not require such expensive catalysts, as do low-temperature fuel cells and even PAFCs. It also offers great flexibility in the use of available fuels (i.e., through in situ reforming of fuels). It has been estimated that the MCFC can achieve an energy conversion efficiency of 52%–60% (LHV) from chemical energy to electrical energy with internal reforming and natural gas as the primary fuel. Some studies have indicated that the MCFC efficiency of methane-to-electricity conversion is the highest attainable by any fuel cell or other single pass/simple cycle generation scheme.

Basic Operating Principle

Fig. 1 illustrates a schematic of a MCFC. A MCFC consists of two porous gas-diffusion electrodes (anode and cathode) and a carbonate electrolyte in liquid form. The electrochemical reaction occurring at the anode and the

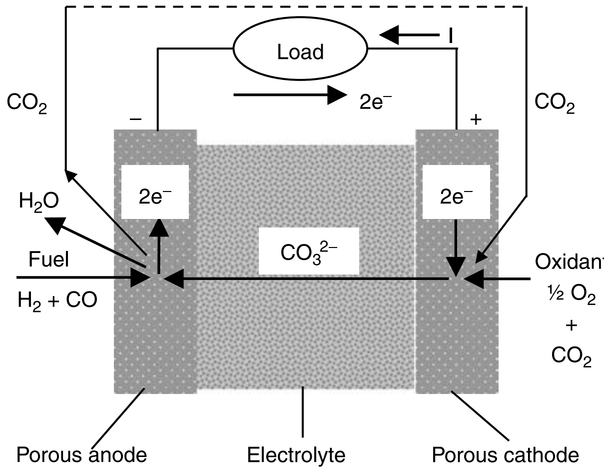


Fig. 1 Schematic of a molten carbonate fuel cell.

cathode is

$$\text{Anode}: \quad H_2 + CO_3^{2-} = H_2O + CO_2 + 2e^-$$

$$\text{Cathode}: \quad \frac{1}{2}O_2 + CO_2 + 2e^- = CO_3^{2-}$$

and the net cell reaction is

$$H_2 + \frac{1}{2}O_2 = H_2O.$$

Besides the hydrogen oxidation reaction at the anode, other fuel gases (such as carbon monoxide, methane, and higher hydrocarbons) are also oxidized by conversion to hydrogen. Although direct electrochemical oxidation of carbon monoxide is possible, it occurs very slowly compared to that of hydrogen. Therefore, the oxidation of carbon monoxide is mainly via the water–gas shift reaction

$$CO + H_2O = CO_2 + H_2$$

which, at the operation temperature of the MCFC, equilibrates very rapidly at catalysts such as nickel. Therefore, carbon monoxide becomes a fuel instead of a contaminant, as in the previously described low-temperature fuel cells. Direct electrochemical reaction of methane appears to be negligible. Hence, methane and other hydrocarbons must be steam-reformed, which can be done either in a separate reformer (external reforming) or in the MCFC itself (the so-called internal reforming).

As a result, water and carbon dioxide are important components of the feed gases to the MCFCs. Water, produced by the main anode reaction, helps shift the equilibrium reactions to produce more hydrogen for the anodic electrochemical reaction. Water must also be present in the feed gas, especially in low-Btu (i.e., high carbon monoxide content) fuel mixtures, to prevent carbon deposition in the fuel gas flow channels supplying the cell or even inside the cell itself. Carbon dioxide, from the fuel exhaust gas, is usually recycled to the cathode as it is required for the reduction of oxygen.

The MCFCs use a molten alkali carbonate mixture as the electrolyte, which is immobilized in a porous lithium aluminate matrix. The conducting species is carbonate ions. Lithiated nickel oxide is the current material of choice for the cathode, and nickel, cobalt, and copper are currently used as anode materials, often in the form of powdered alloys and composites with oxides. As a porous metal structure, it is subject to sintering and creeping under the compressive force necessary for stack operation. Additives such as chromium or aluminum form dispersed oxides and thereby increase the long-term stability of the anode with respect to sintering and creeping. Molten carbonate fuel cells normally have 75%–80% fuel (hydrogen) utilization.

Acceptable Contamination Levels

Molten carbonate fuel cells do not suffer from carbon monoxide poisoning—in fact, they can utilize carbon monoxide in the anode gas as the fuel. However, they are extremely sensitive to the presence of sulfur (<1 ppm) in the reformed fuel (as hydrogen sulfide, H_2S) and oxidant gas stream (SO_2 in the recycled anode exhaust). The presence of HCl, HF, HBr, etc. causes corrosion, while trace metals can spoil the electrodes. The presence of particulates of coal/fine ash in the reformed fuel can clog the gas passages.

Major Technological Problems

The main research efforts involving the MCFCs are focused on increasing lifetime and endurance, and reducing long-term performance decay. The main determining factors for the MCFCs are electrolyte loss, cathode dissolution, electrode creepage and sintering, separator plate corrosion, and catalyst poisoning for internal reforming.

Electrolyte loss results in increased ohmic resistance and activation polarization, and it is the most important and continuously active factor in causing long-term performance degradation. It is primarily a result of electrolyte consumption by the corrosion/dissolution processes of cell components, electric potential-driven electrolyte migration, and electrolyte vaporization. Electrolyte evaporation (usually Li_2CO_3 and/or K_2CO_3) occurs either directly as carbonate or indirectly as hydroxide.

The cathode consists of NiO, which slowly dissolves in the electrolyte during operation. It is then transported toward the anode and precipitates in the electrolyte matrix as Ni. These processes lead to a gradual degradation of cathode performance and shorting of the electrolyte matrix. The time at which shorting occurs depends not only on the CO_2 partial pressure and the cell temperature, but also the matrix structure (i.e., porosity; pore size; and, in particular, thickness of the matrix). Experience indicates that this cell shorting mechanism tends to limit stack life to about 25,000 h under the atmospheric reference gas conditions and much shorter for real operating conditions.

Electrode, especially anode, creepage and sintering (i.e., a coarsening and compression of electrode particles) result in increased ohmic resistance and electrode polarization. NiO cathodes have quite satisfactory sinter and creepage resistance. Creep resistance of electrodes has an important effect on maintaining low-contact resistance of the cells and stacks. The corrosion of the separator plate depends on many factors, such as the substrate; possible protective layers; composition of the electrolyte; local potential and gas composition; and oxidizing and reducing atmospheres at the cathode and anode, respectively.

Poisoning of the reforming catalyst occurs within direct internal reforming MCFCs. It is caused by the evaporation of the electrolyte from the cell components, condensation on the catalyst (which is the coldest spot in the cell), and liquid creep within the cell.

Technological Status

Molten carbonate fuel cell technology is in the first demonstration phase, under product development with full-scale systems at the 250 kW–2 MW range. The short-term goal is to reach a lifetime of 25,000 h, and the ultimate target is 40,000 h. It is estimated that the capital cost is between $1000 and $1600 U.S./kW for the MCFC power systems. The cost breakdown is, at full-scale production levels, about one-third for the stack and two-thirds for the balance of the plant. It is also generally accepted that the cost of raw materials will constitute about 80% of total stack costs. Although substantial development efforts supported by fundamental research are still needed, the available knowledge and number of alternatives will probably make it possible to produce precommercial units in the earlier part of the coming decade at a capital cost of $2000–$4000 U.S./kW. Precompetitive commercial units may be expected some years later, by which time further cost reduction to full competitiveness will be guided by extensive operating experience and increased volume production.

Applications

The MCFCs are being developed for their potential as baseload utility generators. However, they are best applied to distributed power generation and cogeneration for capacities less than 20 MW in size, and in this size range, MCFCs are 50%–100% more efficient than turbines—the conventional power generator. Other applications are foreseen, including pipeline compressor stations, commercial buildings, industrial sites in the near future, and repowering applications in the longer term. Due to its high operation temperature, it has very limited potential for transportation applications. This is because of its relatively low power density and long startup times. However, it may be suitable as a powertrain for large surface ships and trains.

SOLID OXIDE FUEL CELLS

Introduction

Solid oxide fuel cells have emerged as serious alternative high-temperature fuel cells, and they have often been referred to as third-generation fuel cell technology because their commercialization is expected after PAFCs (the first generation) and MCFCs (the second generation).

The SOFC is an all-solid-state power system, including the electrolyte, and it is operated at a high temperature of around 1000°C for adequate ionic and electronic conductivity of various cell components. The all-solid-state cell composition makes the SOFC system simple in concept, design, and construction; two-phase (gas–solid) contact for the reaction zone reduces corrosion and eliminates all the problems associated with the liquid electrolyte management of MCFCs. The high-temperature operation results in fast electrochemical kinetics (i.e., low activation polarization) and no need for noble metal catalysts. The fuel may be gaseous hydrogen, an H_2/CO mixture, or hydrocarbons because the high-temperature operation makes the internal in situ reforming of hydrocarbons with water vapor possible. It is important to note that in SOFCs, carbon monoxide is no longer a contaminant; rather, it becomes a fuel. Even with external reforming, the SOFC fuel feedstock stream does not require extensive steam reforming with shift conversion, as it does for the low-temperature fuel cell systems. More important, the SOFC provides high-quality waste heat, which can be utilized for cogeneration applications or CC operation for additional electric power generation. The SOFC operating condition is also compatible with the coal gasification process, which makes SOFC systems highly efficient when using coal as the primary fuel. It has been estimated that the chemical-to-electrical energy conversion efficiency is at least 50%–60% (LHV), and some estimates go as high as 70%–80% (LHV). Additionally, nitrogen oxides are not produced, and the amount of carbon dioxide released per kilowatt hour is around 50% less than for power sources based on combustion because of the high efficiency of the SOFC.

Basic Operating Principle

As mentioned earlier, both hydrogen and carbon monoxide can be oxidized in the SOFCs directly. Hence, if hydrogen or a hydrogen-rich gas mixture is used as fuel, and oxygen (or air) is used as oxidant, the half-cell reaction becomes (Fig. 2)

$$\text{Anode}: \quad H_2(g) + O^{2-} = H_2O(g) + 2e^-$$

$$\text{Cathode}: \quad 2e^- + \frac{1}{2}O_2(g) = O^{2-}$$

and the overall cell reaction becomes

$$H_2(g) + \frac{1}{2}O_2(g) = H_2O(g).$$

However, if carbon monoxide is provided to the anode instead of hydrogen, the anode reaction becomes

$$\text{Anode}: \quad CO(g) + O^{2-} = CO_2(g) + 2e^-$$

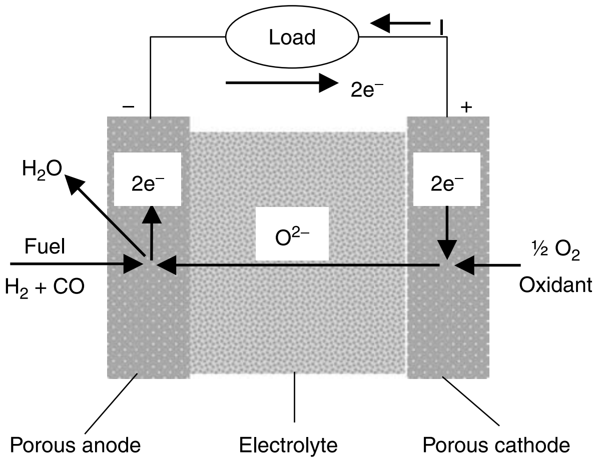


Fig. 2 Schematic of a solid oxide fuel cell (SOFC).

With the cathode reaction remaining the same, the cell reaction becomes

$$CO(g) + \frac{1}{2}O_2(g) = CO_2(g).$$

If the fuel stream contains both hydrogen and carbon monoxide, as is the case for hydrocarbon reformed gas mixtures—especially from the gasification of coal—the oxidation of hydrogen and carbon monoxide occurs simultaneously at the anode, and the combined anode reaction becomes

$$\text{Anode}: \quad aH_2(g) + bCO + (a + b)O^{2-} = aH_2O(g) + bCO_2 + 2(a + b)e^-$$

Consequently, the corresponding cathode and overall cell reaction become

$$\text{Cathode}: \frac{1}{2}(a + b)O_2 + 2(a + b)e^- = (a + b)O^{2-}$$

$$\text{Cell}: \quad \frac{1}{2}(a + b)O_2 + aH_2 + bCO = aH_2O + bCO_2$$

The solid electrolyte in SOFCs is usually yttria-stabilized zirconia (YSZ); thus, a high operating temperature of around 1000°C is required to ensure adequate ionic conductivity and low ohmic resistance. This is especially important because the cell open-circuit voltage is low as compared with low-temperature fuel cells—typically between 0.9 and 1 V under the typical working conditions of the SOFCs. The high-temperature operation of the SOFCs makes the activation polarization very small, resulting in the design operation in the range dominated by the ohmic polarization. The conventional material for the anode is nickel–YSZ-cermet, and the cathode is usually made of lanthanum–strontium-manganite. Metallic current collector plates of a high-temperature, corrosion-resistant, chromium-based alloy are typically used. Recently, low-temperature SOFCs are under intensive R&D, with operating temperatures ranging from 600 to 800°C. At such lower temperatures, cheaper materials are expected to be used for the cell components, including interconnectors. However, low-temperature SOFCs are still in their early stage of development, and significant technical issues need to be resolved before precommercial demonstration of the technology.

Acceptable Contamination Levels

Because of their high temperatures, the SOFCs can better tolerate impurities in the incoming fuel stream. They can operate equally well on dry or humidified hydrogen or carbon monoxide fuel, or on mixtures of them. But hydrogen sulfide (H_2S), hydrogen chloride (HCl), and ammonia (NH_3) are impurities typically found in coal-gasified products, and each of these substances is potentially harmful to the performance of SOFCs. The main poisoning factor for SOFCs is H_2S. Though the sulfur tolerance level is approximately two orders of magnitude greater than other fuel cells, the level is below 80 ppm. However, studies have shown that the effect of hydrogen sulfide (H_2S) is reversible, meaning that the cell performance will recover if hydrogen sulfide is removed from the fuel stream or clean fuel is provided after the contaminant poison has occurred.

Major Technological Problems

The high-temperature operation of the SOFCs places stringent requirements on materials used for cell construction, and appropriate materials for cell components are very scarce. Therefore, the key technical challenges are the development of suitable materials and the fabrication techniques. Of the material requirements, the most important consideration is the matching of the thermal expansion coefficients of electrode materials with those of the electrolyte to prevent cracking or delamination of SOFC components either during high-temperature operation or heating/cooling cycles. One of the remedies for the thermal expansion mismatch is to increase the mechanical toughness of the cell materials by developing new materials or by doping the existing materials with SrO and CaO.

The electrode voltage losses are reduced when the electrode material possesses both ionic and electronic conductivities (the so-called mixed conduction), for which the electrochemical reactions occur throughout the entire surface of the electrode rather than only at the three-phase interface of, for example, the cathode, the air (gas phase), and the electrolyte. Therefore, it is important for performance enhancement to develop mixed-conduction

Table 1 Operational characteristics and technological status of intermediate and high temperature fuel cells

Type of fuel cells	Operating temperature (°C)	Power density (mW/cm^2) (present) projected	Projected rated power level (kW)	Fuel efficiency	Lifetime projected (h)	Capital cost projected (U.S.$/kW)	Application areas
PAFC	160–220	(200) 250	100–5,000	55	>40,000	3,000	Dispersed & distributed power
MCFC	600–700	(100) >200	1,000–100,000	60–65	>40,000	1,000	Distributed power generation
SOFC	800–1,000	(240) 300	100–100,000	55–65	>40,000	1,500	Baseload power generation; auxiliary power for large transport vehicles

materials for both the cathode and anode, which have a good thermal expansion match with the electrolyte used and good electrical conductivity to reduce the ohmic polarization that dominates the SOFC voltage losses.

Another focus of current development is the intermediate-temperature SOFCs, operating at around 800°C for better matching with the bottoming turbine cycles and lessening requirements for the cell component materials. Again, appropriate materials with adequate electrical conductivity are the key areas of the development effort, and thermal expansion matching among the cell components is still necessary.

Technological Status

There are three major configurations for SOFCs: tubular, flat plate, and monolithic. Even though SOFC technology is in the developmental stage, the tubular design has gone through development at Westinghouse Electric Corporation since the late 1950s, and it is now being demonstrated at user sites in a complete operating fuel cell power unit of nominal 25 kW (40 kW maximum) capacity. The flat plate and the monolithic designs are at a much earlier development status, typified by subscale, single-cell, and short stack development (up to 40 cells). The present estimated capital cost is $1500 U.S./kW but is expected to be reduced with improvements in technology. Therefore, the SOFCs may become very competitive with the existing technology for electric power generation. However, it is believed that the SOFC technology is at least 5–10 years away from commercialization.

Applications

Solid oxide fuel cells are very attractive in electrical utility and industrial applications. The high operating temperature allows them to use hydrogen and carbon monoxide from natural gas steam reformers and coal gasification plants—a major advantage as far as fuel selection is concerned. Solid oxide fuel cells are being developed for large (>10 MW, especially 100–300 MW) baseload stationary power plants with coal as the primary fuel. This is one of the most lucrative markets for this type of fuel cell.

A promising field for SOFCs is the decentralized power supply in the MW range, where the SOFC gains interest due to its capability to convert natural gas without external reforming. In the capacity range of one MW to some tens of MWs, the predicted benefits in electrical efficiency of SOFC-based power plants over conventional methods of electricity generation from natural gas can only be achieved by an internal-reforming SOFC. So internal reforming is a major target of present worldwide SOFC development.

Another application that is gaining momentum is as auxiliary power for large transport vehicles (e.g., heavy-duty transport trucks), although this application is still in its early stage.

CONCLUDING REMARKS

A summary of the preceding description of the three major types of intermediate- and high-temperature fuel cells—namely, the phosphoric acid, molten carbonate, and SOFCs—is provided in Table 1, including operational characteristics and technological status.

Phosphoric acid fuel cells are the most commercially developed fuel cells operating at intermediate temperatures. Phosphoric acid fuel cells are being applied for distributed or on-site cogeneration in such locations as hospitals, hotels, offices, schools, etc. with relative high conversion efficiency. The high-temperature fuel cells, like MCFCs and SOFCs, may be most appropriate for cogeneration and combined cycle systems (with gas or steam turbines as the bottoming cycle). Molten carbonate fuel cells have the highest energy efficiency attainable from methane-to-electricity conversion (in the size range of 250 kW–20 MW), whereas SOFCs are best suited for baseload utility applications operating on coal-derived gases. A recent interest in SOFCs is as auxiliary power for large transport vehicles. It is estimated that MCFC technology is 5–10 years away from commercialization, and SOFC development will likely follow even later.

BIBLIOGRAPHY

1. Blomen, L.J.M.J.; Mugerwa, M.N. *Fuel Cell Systems*; Plenum Press: New York, 1993.
2. Larminie, J.; Dicks, A. *Fuel Cell Systems Explained*, 2nd Ed.; Wiley: United Kingdom, 2003.
3. Li, X. *Principles of Fuel Cells*; Taylor & Francis: New York, 2006.
4. Kordesch, K.; Simader, G. *Fuel Cells and Their Applications*; VCH Press: New York, 1996.
5. Appleby, A.J.; Foulkes, F.R. *Fuel Cell Handbook*; Van Nostrand Reinhold Company: New York, 1989.
6. Hoogers, G. *Fuel Cell Technology Handbook*; CRC Press: Boca Raton, FL, 2002.
7. Vielstich, W.; Gasteiger, H.; Lamm, A. *Handbook of Fuel Cells—Fundamentals, Technology, Applications*; Wiley: United Kingdom, 2003.

Fuel Cells: Low Temperature

Xianguo Li
Department of Mechanical Engineering, University of Waterloo, Waterloo, Ontario, Canada

Abstract
This entry provides an overview of low-temperature fuel cells, including alkaline fuel cells, proton exchange membrane fuel cells, and direct methanol fuel cells. For each of them, a brief description is given of their operational characteristics, acceptable contamination levels, major technical barriers to commercialization, technological status, economics, and major intended applications.

INTRODUCTION

Fuel cell technology has improved dramatically over the past few years, once again capturing public and industry attention concerning the prospect of fuel cells as practical power sources for terrestrial applications. The fuel cell is a highly energy-efficient and environmentally friendly technology for power generation that is also compatible with alternative fuels, renewable energy sources, and carriers for sustainable development.

Fuel cells offer additional advantages for both mobile and stationary power generation applications. The fuel cell, as an electrochemical device, has no moving components except for peripheral equipment. As a result, its operation is very quiet, virtually without vibration and noise. It is capable of on-site cogeneration with no need for long-distance power transmission lines that consume approximately 10% of electric energy delivered to consumers in North America. Its inherent modularity allows for simple construction and operation. It has possible applications in baseload electricity generation as well as in dispersed, distributed, and portable power generation because it may be made in any size raging from a few watts up to a megawatt-scale plant with equal efficiency. Its fast response to changing load conditions while maintaining high efficiency makes it ideally suited to load-following applications. Its high efficiency results in lower chemical, thermal, and carbon dioxide emissions for a given amount of power generation.

At present, fuel cell technology is being used routinely in many specific areas, notably in space explorations, where fuel cells operate on pure hydrogen and oxygen with over 70% [lower heating value (LHV)] efficiency and produce potable water as the only byproduct. There are now approximately 200 fuel cell units (for terrestrial applications) operating in 15 countries. Technical progress has been impressive in terms of higher power density, better performance, reduced capital, and reduced maintenance and operation cost. Competitively priced fuel cell-based power generation systems with advanced features are being developed, including utility power plants and zero-emission vehicles. In light of decreasing fossil fuel reserves and increasing energy demands worldwide, fuel cells will probably become one of the major energy technologies with strong international competition in the 21st century.

Outside of the space industry, the major commercial applications of fuel cells include electric power generation in the utility industry and a zero-emission powertrain in the transportation sector. For these practical applications, the efficiencies of fuel cells range somewhere from 40 to 65%, based on the LHV of hydrogen. Typically, a single cell's electric potential is only about 1 V, and this decreases due to various loss mechanisms under operational conditions. Thus, multiple cells must be connected in series to achieve a useful voltage. These connected cells are often referred to as a fuel cell stack. A fuel cell system consists of one or more fuel cell stacks connected in series or parallel, and the auxiliaries whose composition depends on the type of fuel cells and the kind of primary fuels used. The major accessories include a thermal management (or cooling) subsystem, a fuel supply, a storage and processing subsystem, and an oxidant (typically air) supply, and conditioning subsystem.

There are six major types of fuel cells that have been under research, development, and demonstration (RD&D) for practical applications. They are:

- Alkaline fuel cells (AFCs)
- Proton exchange membrane fuel cells (PEMFCs)
- Direct methanol fuel cells (DMFCs)
- Phosphoric acid fuel cells (PAFCs)
- Molten carbonate fuel cells (MCFCs)
- Solid oxide fuel cells (SOFCs)

A critical parameter that determines the potential applications of each type of fuel cell is the operating temperature. For instance, AFCs, DMFCs, and PEMFCs

Keywords: Fuel cells; Alkaline; Proton exchange membrane; Direct methanol.

have potential applications in transportation because of their low operating temperatures. Consequently, they have a very short startup period, consisting of only a few minutes. These types of fuel cells are discussed in the first entry, under the title "Low Temperature Fuel Cells." Meanwhile, phosphoric acid, molten carbonate, and SOFCs producing high-temperature heat are more complex to run and are better suited for stationary applications like combined heat and power generation (CHP). The fuel cells in this category will be discussed in the second entry, titled "Intermediate and High Temperature Fuel Cells."

The following sections describe each of the three low-temperature fuel cells. Their technological status is presented with respect to their basic operating principles, acceptable contamination levels, state-of-the-art technology, major technical barriers to commercialization, economics, and suitability for utility and transportation applications.

ALKALINE FUEL CELLS

Alkaline fuel cells are among the first fuel cells to have been studied and developed for practical applications, and they are the first fuel cells to have been applied successfully to routine applications, including space-shuttle missions in the United States and similar space exploration endeavors in China and Europe, where pure hydrogen and oxygen are used as reactants. Because of their success in the space programs, AFCs are also the type of fuel cells studied in the largest number of fuel cell development programs for terrestrial applications, particularly in Europe. However, almost all of the AFC development programs have now ended, with only a few remaining activities related to the AFC RD&D still taking place in Europe.

Alkaline fuel cells have the highest energy conversion efficiency among all types of fuel cells under the same operating conditions if pure hydrogen and pure oxygen are used as reactants. This is one reason why AFCs were selected for the U.S. space-shuttle missions. The AFCs used in the shuttle missions are operated at about 200°C for maximum performance (i.e., high energy conversion efficiency of over 70% and high power density, both of which are critical for space applications), and the alkaline electrolyte is potassium hydroxide (KOH) solution immobilized in an asbestos matrix. The AFCs must operate at high pressure to prevent boiling and depletion of the liquid electrolyte. These severe operating conditions of high temperature and high pressure dictate an extremely stringent requirement for cell component materials because they must withstand the extreme corrosive oxidizing and reducing environment of the cathode and anode. To meet these space-program requirements, precious metals such as platinum, gold, and silver are used for the electrodes, but these precious metals need not be present for the electrochemical reactions leading to electric power generation. Each shuttle flight contains a 36 kW AFC power system. Its purchase price is about $28.5 million U.S., and it costs NASA an additional $12 million to $19 million U.S. annually for operation and maintenance. Although the manufacturer claims about 2400 h of lifetime, NASA's experience indicates that the real lifetime is only about 1200 h. With sufficient technology development, 10,000 h are expected to be the life potential (or upper limit) for the AFC system in the coming years—a prediction based on the nature of the AFC systems and the data accumulated on both stacks and single cells.

The typical working temperature of AFC power systems for commercial and terrestrial applications ranges from 20 to 90°C, and the electrolyte is a KOH solution (30%–35%). There are four different AFC cell types investigated:

- Cell with a free liquid electrolyte between two porous electrodes
- ELOFLUX cell with liquid KOH in the pore systems
- Matrix cell with the electrolyte is fixed in the electrode matrix
- Falling film cell

Research groups worked on the development of technical AFC systems, although most of these groups have now terminated their research. The group names and the current status of their work are listed below:

- Siemens, Erlangen, Germany: stopped working on the AFCs
- VARTA AG, Kelkheim, Germany: terminated working on preparation of technical electrodes and technical fuel stacks in 1993
- GH, Kassel, Germany: stopped working on preparation of technical electrodes and technical fuel cell stacks in 1994
- ISET, Kassel, Germany: stopped working on technical AFC systems in 1994
- DLR-ITT (German Aerospace Research Establishment), Stuttgart, Germany: stopped working on the investigation of the degradation of technical electrodes, the development of new catalysts for AFCs, and the theoretical simulation of stacks and systems in 1994
- ELENCO, Antwerpen, Belgium: stopped working on electrodes, stacks, and systems, as well as the bus demonstration program (went bankrupt in 1995)
- Royal Institute of Technology, Stockholm, Sweden: working on the field of stationary fuel cells powered by biofuels
- Hoechst AG, Frankfurt, Germany: terminated working on AFC electrodes and stopped development of the falling film cell

- Technical University, Graz, Austria: planning an investigation of degradation effects
- Zevco, London, England: demonstration of AFCs for mobile applications, particularly for bus with ELENCO's technology

The following is a list of some selected technical applications and demonstration projects for AFCs in Europe in recent years:

- Space applications
  - AFC system for the European space shuttle HERMES (Siemens, ELENCO, VARTA/GHK)
  - Bipolar matrix AFC system Photon (Ural Electrochemical Integrated Plant)
- Vehicles
  - Forklift truck, VARTA AFC
  - Volkswagon (VW) van, 14 kW ELENCO AFC and battery
  - VW van, 17.5 kW Siemens AFC
  - Submarine, 100 kW Siemens AFC
  - EUREKA-Bus, 80 kW ELENCO AFC and battery
- Decentralized energy supply
  - Meteorological station, 5 kW, VARTA AFC for long-term operation
  - TV-transmitting installation Ruppertshain, 100 kW VARTA AFC
  - Mobile current supply unit for the Belgian Geological Service, 40 kW ELENCO AFC
- Energy storage
  - Solar hydrogen-demonstration plant in Neunburg vorm Wald, 6.5 kW Siemens AFC
  - Solar hydrogen system at Fachhochschule Wiesbaden, 1.2 kW ELENCO AFC

Many problems concerning AFCs are described in the literature. The most important include:

- *Method of preparing the electrode*. The electrodes consist of porous material covered with a layer of catalyst. In general, it is very difficult to distribute the catalyst at the surface and produce a defined pore system for transporting the reactants.
- *Costs of the electrode, stacks, and fuel cell systems*. The preparation of electrodes with noble metal catalysts is very expensive. In general, the electrodes are manufactured in a small-scale production with high overhead costs.
- *Lifetime of the electrode/degradation*. The electrolyte is very corrosive, and the catalyst materials are sensitive to high polarization. Using nickel and silver as catalysts to reduce the costs of the fuel cell leads to a high degradation of these catalysts.
- *Diaphragm made of asbestos*. The diaphragm of low-temperature AFCs is made of asbestos, but this material is a health hazard, and in some countries its use is even banned. Therefore, new diaphragms should be developed, but it is difficult to find a material with similar behavior in an alkaline electrolyte.
- *Carbon dioxide-contaminated fuel and oxidant streams resulting in carbonating of electrolyte and electrodes*. The electrolyte's intolerance of carbon dioxide is the most important disadvantage of air-breathing AFCs with reformulated gases from primary fossil fuels.

Other problems associated with AFC power systems relate to the safety and reliability of AFC power systems. For example, the liquid KOH electrolyte contained in an asbestos matrix can withstand only a 5 psi limit of pressure differential between the anode and cathode reactant gases. This dictates the need for sophisticated pinpoint pressure control during operation, including the transient, startup, and shutdown processes. There is also a safety issue due to the great likelihood of mixing the reactants in the AFC system, with the potential result of a serious fire. In terms of general safety considerations, the use of the corrosive potassium hydroxide electrolyte in the AFCs requires hazardous materials, and the handling of the asbestos matrix poses potential hazards to one's health. With flowing reactant gases, the potential for the gradual loss of the liquid electrolyte, drying of the electrolyte matrix, reactant crossover of the matrix, and ensuing life-limiting reactant mixing (or actual AFC stack failure due to fire) is very real in the AFC system.

The major technical challenge is that alkaline electrolytes, like potassium or sodium hydroxide, are adversely affected by 300–350 ppm of carbon dioxide. Atmospheric air has carbon dioxide concentration in both cathode and anode gases greater than 10–100 ppm by volume. Most applications require the use of atmospheric air as the oxidant due to technical and economic considerations. For municipal electric applications, hydrocarbon fuels, especially natural gas, are expected to be the primary fuel. When they are reformed into hydrogen-rich gases, they invariably contain a significant amount of carbon dioxide. For example, steam-reforming of natural gas results in a reformate gas consisting of approximately 80% hydrogen; 20% carbon dioxide; and a trace amount of other components, such as carbon monoxide. In addition, the carbonaceous products of aging and corrosion shorten AFC life because they degrade the alkaline electrolyte. Whether they originate as impurities in the gaseous reactants or from some fuel cell materials, oxides of carbon will chemically react with the alkaline electrolyte and produce irreversible decay, subsequently decreasing performance and shortening life span. As a result, AFCs are currently restricted to specialized applications where pure hydrogen and oxygen are utilized. The revival of AFC technology will depend almost completely on successfully solving the "CO_2 syndrome"—i.e., efficiently and economically scrubbing carbon dioxide.

Foss–Gre

Claims have been made of successful resolution of carbon dioxide poisoning in the AFCs, especially in utility applications. Some optimistic estimates indicate that AFC stack costs are similar to all other low-temperature fuel cell systems. Production costs for the AFC systems seem to be the lowest. A price of approximately $400–$500/kW U.S. has been quoted by using present-day technologies and knowledge in large-scale production. However, small-scale commercial production cost is estimated to be five to ten times higher.

PROTON EXCHANGE MEMBRANE FUEL CELLS

Introduction

The PEMFC is also called the solid polymer (electrolyte) fuel cell. It is perhaps the most elegant of all fuel cell systems in terms of design and mode of operation. It was the first type of fuel cell put into practical application in the Gemini space missions between 1962 and 1966. It consists of a solid polymeric membrane acting as the electrolyte. The solid membrane is an excellent proton conductor, sandwiched between two platinum-catalyzed porous carbon electrodes. It has fast start capability and yields the highest output power density among all types of fuel cells. Because of the solid membrane as electrolyte, there is no corrosive fluid spillage hazard, and there is a lower sensitivity to orientation. The electrolyte is not volatile, and there are minimal corrosion concerns. It has zero pollutant emissions, with only potable water as a byproduct when hydrogen is used as fuel. As a result, the PEMFC is particularly suited for vehicular power application, although it is also being considered for stationary and portable power application (albeit to a lesser degree).

The proton-conducting polymer membrane belongs to a class of materials known as ionomers or polyelectrolytes, which contain functional groups that will dissociate in the presence of water. The dissociation produces ions fixed to the polymer and simple counter-ions, which can exchange freely with ions of the same sign from the solution. The polyelectrolytes currently available feature cations as the counterions. For hydrogen, the cation is a proton.

The membrane must be fully hydrated to have adequate ion conductivity. As a result, the fuel cell must be operated under conditions where the byproduct water does not evaporate faster than it is produced, and the reactant gases, both hydrogen and oxygen, need to be humidified. Managing water and temperature in the membrane is critical for efficient cell performance; dynamic controls are needed to match the varying operating conditions of the fuel cell. Because of the limitations imposed by the membrane and problems with the water balance, the operating temperature of PEMFCs is usually less than 120°C—typically 80°C. This rather low operating temperature requires the use of noble metals as catalysts in both the anode and cathode side, with generally higher catalyst loadings than those used in PAFCs.

Currently, the polymer electrolyte is made of a perfluorinated sulfonic acid membrane; it is essentially acid, though in solid polymeric form. Hence, a PEMFC is essentially an acid electrolyte fuel cell, with its operational principle nearly the same as a PAFC. As a result, most PEMFC design, material selection, component fabrication, etc. are similar to those of PAFCs. The only difference is the humidification of reactant gases dictated by the membrane performance. Reactant humidification is often achieved by a number of techniques—e.g., by passing the gas stream through a water column, by using in-stack humidification sections of the cell and membrane arrangement, and by spraying water into the reactant streams. In the early stages of PEMFC development, the membranes were based on polystyrene, but since 1968, a Teflon-based product known as Nafion by DuPont has been used. This offers high stability, high oxygen solubility, and high mechanical strength.

Basic Operating Principle

The schematic of a single PEM fuel cell is illustrated in Fig. 1. The PEMFC requires hydrogen gas as the fuel and oxygen (typically air) as the oxidant. The half-cell reactions are

$$\text{Anode}: \quad H_2 = 2H^+ + 2e^-$$

$$\text{Cathode}: \quad \tfrac{1}{2}O_2 + 2H^+ + 2e^- = H_2O$$

and the overall cell reaction is

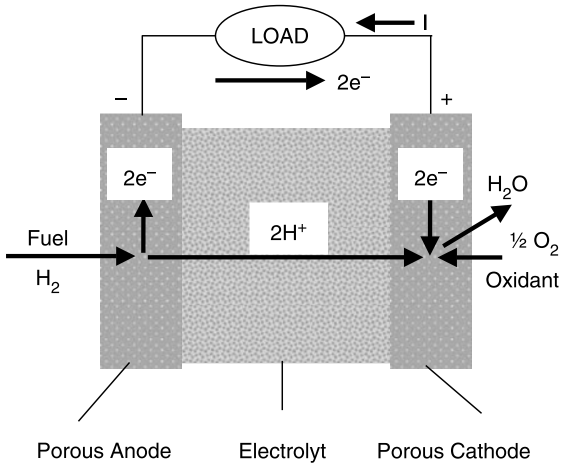


Fig. 1 Schematic of a PEMFC.

$$H_2 + \tfrac{1}{2}O_2 = H_2O + \text{Heat Generated} + \text{Electric Energy}.$$

The current PEMFCs use perfluorinated sulfonic acid membrane (almost exclusively Nafion) as the proton-conducting electrolyte; carbon paper or cloth as the anode and cathode backing layers; and platinum or its alloys, either pure or supported on carbon black, as the catalyst. The bipolar plate with the reactant gas flow fields is often made of graphite plate. The stoichiometry is between 1.1 and 1.2 for the fuel, and 2 for the oxidant (oxygen). The PEMFCs usually operate at about 80°C and 1–8 atm pressure. The pressures, in general, are maintained equally on either side of the membrane. Operation at high pressure is necessary to attain high power densities, particularly when air is chosen as the cathodic reactant.

To prevent the membrane dryout from leading to local hot spots, crack formation, performance degradation, and lifetime reduction, both fuel and oxidant streams are fully humidified, and the operating temperature is limited by the saturation temperature of water corresponding to the operating pressure. The liquid water formed at the cathode does not dissolve in the electrolyte membrane and is usually removed from the cell by the excess oxidant gas stream. The accumulation of liquid water in the cathode backing layer blocks the oxygen transfer to the catalytic sites, resulting in the phenomenon called water flooding, which causes performance reduction. Local hot and cold spots will cause the evaporation and condensation of water. Thus, an integrated approach to thermal and water management is critical to PEMFCs' operation and performance, and a proper design must be implemented.

Acceptable Contamination Levels

As an acid electrolyte fuel cell operating at low temperature, the PEMFC is primarily vulnerable to carbon monoxide poisoning. Even a trace amount of carbon monoxide (CO) drastically reduces the performance levels, although the CO poisoning effect is reversible and does not cause permanent damage to the PEMFC system. Further, performance reduction due to CO poisoning can take up to two hours to reach steady state. This transient effect may have profound implications for transportation applications. Therefore, the PEMFC requires the use of a fuel virtually free of CO (it must be less than a few parts per million). Additionally, high-quality water free of metal ions should be used for the cell cooling and reactant humidification to prevent the contamination of the membrane electrolyte. This requirement has a severe implication on the materials that can be used for cell components. On the other hand, carbon dioxide does not affect PEMFC operation and performance except through the effect of reactant dilution (the Nernst loss). However, carbon dioxide can produce carbon monoxide in the presence of water vapor available in the PEMFCs via reverse water–gas shift reaction. PEMFCs are also susceptible to the poisoning of sulfur and its compounds—hydrogen sulfide in particular—although this is less quantified than the CO poisoning at this stage of development.

Major Technological Problems

For practical applications, PEMFC performance in terms of energy efficiency, power density (both size and weight), and capital cost must be improved further. This can be accomplished by systematic research in:

- New oxygen reduction electrocatalysts. This includes the reduction of precious-metal platinum and its alloys loading from 4 to 0.4 mg/cm^2 or lower without affecting the long-term performance and the lifetime, and the development of CO-tolerant catalysts.
- New types of polymer electrolyte with higher oxygen solubility, thermal stability, long life, and low cost. A self-humidified membrane or a polymer without the need of humidification will be ideal for PEMFC operation and performance enhancement with significant simplification of system complexities and reduction of cost.
- Profound changes in oxygen (air) diffusion electrode structure to minimize all transport-related losses. The minimization of all transport losses is the most promising direction for PEMFC performance improvement.
- Optimal thermal and water management throughout the individual cells and the whole stack to prevent local hot and dry spot formation and to prevent water flooding of the electrode.

Fig. 2 illustrates the schematic of a PEMFC stack and its major components. In addition to the above issues, the development of low-cost lightweight materials for construction of reactant gas flow fields and bipolar plates is one of the major barriers to the large-scale commercialization of PEMFCs. A successful solution to this problem will further increase the output power density, and such a solution includes an optimal design of flow fields with the operating conditions and an appropriate selection of materials and fabrication techniques. As much as a 20% improvement in the performance of PEMFC stacks can be obtained just through the appropriate design of flow channels. The current leading technologies for bipolar plate design include injection-molded carbon–polymer composites, injection-molded and carbonized amorphous carbon, assembled three-piece metallic, and stamped unitized metallic.

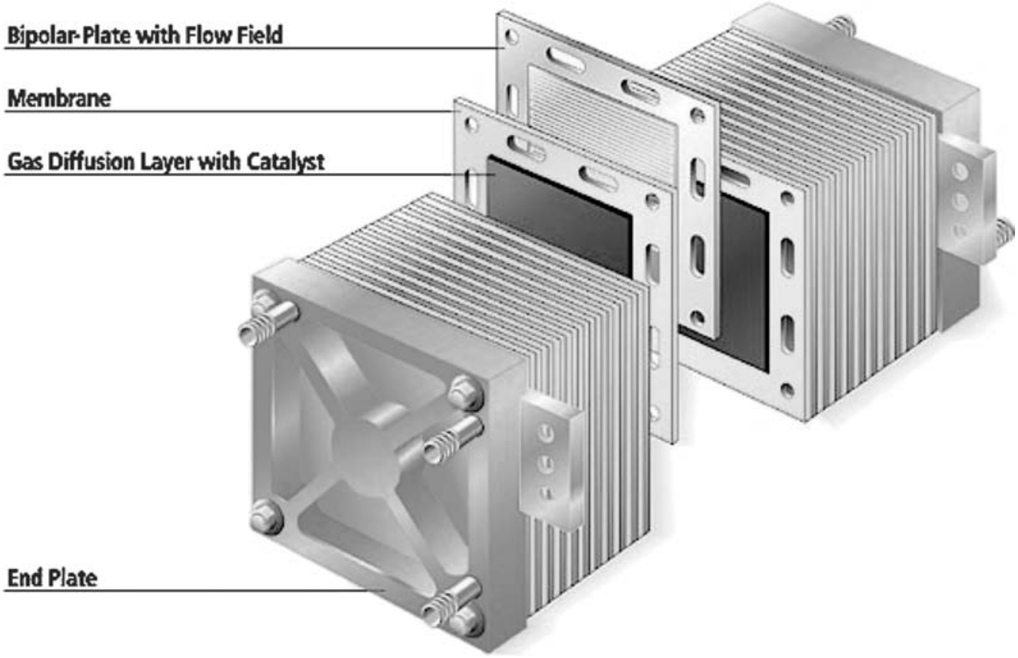


Fig. 2 Schematic of a PEMFC stack and its components.
Source: From Ballard Power Systems; www.ballard.com (see Ref. 3).

Technological Status

Proton exchange membrane fuel cells have achieved a high power density of over 1 kW/kg and 1.0 kW/L for the stack—perhaps the highest among all types of fuel cells currently under development. It is also projected that the power density may be further improved, to between 2 and 3 kW/L, with unitized metallic (stainless steel) bipolar plates. The capital cost varies from the most optimistic estimate of $1500/kW U.S. to the most pessimistic estimate of $50,000/kW U.S. at the current technology, and is projected to reach approximately $30/kW U.S. for net system power, assuming a 10-fold to 20-fold reduction in the membrane and catalyst cost, and also considering large-scale productions. It is expected that PEMFC technology is between five and ten years away from commercialization. Precommercial demonstrations for buses and passenger vehicles are under way with increasing intensity; the first demonstration for residential combined heat and power application began in late 1999. The use of PEMFCs in powering portable and mobile electronics (such as laptops) has also already begun. Tables 1 and 2 show the technical targets set by the U.S. Department of Energy for the year 2010 and 2015, as well as the technical achievements made so far (up to 2005). Table 3 shows the technical progresses up to 2005 and the technical targets for 2007 and 2010 set by Ballard Power Systems, a world leader in PEMFC technology development for transportation applications.

Applications

Proton exchange membrane fuel cells have a high power density, a variable power output, and a short startup time due to low operating temperatures; the solid polymer electrolyte is virtually corrosion free and can withstand a large pressure differential (as high as 750 psi, reported by NASA) between the anode and cathode reactant gas streams. Accordingly, PEMFCs are suitable for use in the transportation sector, and they are currently considered the best choice for zero-emission vehicles. Their high power density and small size make them primary candidates for light-duty vehicles, though they are also used for heavy-duty vehicles. However, conventional gasoline and diesel engines are extremely cheap (estimated to cost between $30 and $50/kW U.S.). Therefore, the cost of PEMFC systems must be lowered by at least two to three orders of magnitude to become competitive with that of conventional heat engines in the transportation arena.

A reformer with carbon monoxide and sulfur cleaning is necessary for electricity generation from hydrocarbon fuels. The cost of the reforming system is estimated to be about the same as the cost of the fuel cell stack itself, as well as that of other ancillary systems. Additionally, the optimal chemical to electric conversion efficiency is approximately 40%–45% (LHV), and the low operating temperature makes the utilization of waste heat difficult or impossible for the reforming of hydrocarbon fuels, cogeneration of heat, and combined cycles. On the other hand, conventional thermal power plants with combined

Table 1 Technical targets for PEM fuel cell power systems operating directly on hydrogen for transportation application (80 kWe net power)[a]

Characteristic	Units	2004 Status	2005 Status	2010	2015
Energy efficiency[b] at 25% of rated power	%	59	60	60	60
Energy efficiency at rated power	%	50	50	50	50
Power density	W/L	450[c]	500	650	650
Specific power	W/kg	420[c]	500	650	650
Cost[d]	US$/kWe	120[e]	125	45	30
Transient response (time from 10 to 90% of rated power)	s	1.5	2	1	1
Cold start-up time to 90% of rated power					
At −20°C ambient temperature	s	120	60	30	30
At +20°C ambient temperature	s	60	30	15	15
Durability with cycling	h	~1000[f]	2000	5000[g]	5000[g]
Survivability[h]	°C	−20	−30	−40	−40

The target for 2015 represents the ultimate target.
[a] Targets exclude hydrogen storage.
[b] Ratio of DC output energy to the LHV of the input fuel (hydrogen). Peak efficiency occurs at about 25% rated power.
[c] Based on corresponding data in Table 2 divided by 3 to account for ancillaries (balance of plant).
[d] Based on 2002 dollars and cost projected to high-volume manufacturing (500,000 stacks per year).
[e] Based on 2004 TIAX Study.
[f] Steady-state durability is 9000 h.
[g] Includes typical drive cycle.
[h] Performance targets must be achieved at the end of 8-hour cold-soak at the temperature given.
Source: From US Department of Energy (see Ref. 1).

gas and steam turbines boast energy efficiency approaching 60% (LHV), with a very low capital cost of $1000/kW U.S. Therefore, the best possible application of PEMFC systems is to use them in remote regions, as well as for combined heat and power application in residential sectors.

In addition, NASA is testing the feasibility of using PEMFC power systems for its space programs (mainly space-shuttle missions) in place of its three current 12 kW AFC power modules. As discussed in the AFCs section, NASA is motivated by the extremely high cost, low lifetime, and maintenance difficulty associated with its present-day AFC systems. Currently, NASA's feasibility study is in its second phase, using parabolic flight tests in airplanes to simulate low-gravity environments. If everything proceeds accordingly, NASA will conduct real-time tests in shuttles within a couple of years.

DIRECT METHANOL FUEL CELLS

Introduction

The fuel cells described above require the use of gaseous hydrogen or liquid/solid hydrocarbon fuels (e.g., methanol reformed to hydrogen). Pure oxygen or oxygen in air is used as the oxidant. Hence, these fuel cells are often referred to as hydrogen–oxygen or hydrogen–air fuel cells. The use of gaseous hydrogen as a fuel presents a number of practical problems, including storage system weight and volume, as well as handling and safety issues, especially for consumer and transportation applications. Although liquid hydrogen has a higher energy density, the liquefaction of hydrogen requires roughly one-third of the specific energy, and the thermal insulation required increases the volume of the reservoir significantly. The use of metal hydrides decreases the specific energy density, and the weight of the reservoir becomes excessive. The size and weight of a power system are both extremely important for transportation applications, as they directly affect fuel economy and vehicle capacity—such factors are less critical for stationary applications. The low volumetric energy density of hydrogen also limits the distance between vehicle refuelings.

Methanol as a fuel offers ease of handling and storage as well as potential infrastructure capability for distribution. Methanol also has a higher theoretical energy density than hydrogen (5 kWh/L compared with 2.6 kWh/L for liquid hydrogen). Easy refueling is another advantage for methanol. However, a reformer is needed in conventional hydrogen–air or hydrogen–oxygen fuel cells, which adds complexity and cost, as well as leading to the

Table 2 Technical targets for PEM fuel cell stacks operating directly on hydrogen for transportation application (80 kWe net power)[a]

Characteristic	Units	2004 Status	2005 Status	2010	2015
Energy efficiency[b] at 25% of rated power	%	65	65	65	65
Energy efficiency[b] at rated power	%	55	55	55	55
Stack power density[c]	W/L	1330[d]	1500	2000	2000
Stack specific power	W/kg	1260[d]	1500	2000	2000
Cost[e]	US$/kWe	75[f]	65	30	20
Transient response (time from 10 to 90% of rated power)	s	1	2	1	1
Cold startup time to 90% of rated power					
At −20°C ambient temperature	s	120	60	30	30
At +20°C ambient temperature	s	<60	30	15	15
Durability with cycling	h	~1000[g]	2000	5000[h]	5000[h]
Survivability[i]	°C	−40	−30	−40	−40
Precious metal loading[j]	g/kW	1.3	2.7	0.3	0.2

The target for 2015 represents the ultimate target.
[a] Targets exclude hydrogen storage and fuel cell ancillaries (thermal, water, and air management subsystems).
[b] Ratio of DC output energy to the LHV of the input fuel (hydrogen) stream. Peak efficiency occurs at about 25% rated power. Assumes system efficiency is 92% of stack efficiency.
[c] Power refers to net power (i.e., stack power minus auxilliary power). Volume is "box" volume, including dead space, and is defined as the water-displaced volume times 1.5 (packaging factor).
[d] Average from Fuel Cells 2000, http://www.fuelcells.org/info/charts.html#fcvs, April 2004.
[e] Based on 2002 dollars and cost projected to high-volume manufacturing (500,000 stacks per year).
[f] Based on 2004 TIAX Study.
[g] Steady-state durability is 9000 h.
[h] Includes typical drive cycle.
[i] Performance targets must be achieved at the end of 8-hour cold-soak at the temperature given.
[j] Equivalent total precious metal loading (anode + cathode): 0.1 mg/cm^2 by 2010 at rated power. Precious metal target based on cost target of <$3/kWe precious metals in MEA [at $450/troy ounce ($15/g), <0.2 g/kWe].
Source: From US Department of Energy (see Ref. 1).

production of undesirable pollutants such as carbon monoxide. The addition of a reformer also increases response time.

Therefore, direct oxidation of methanol is an attractive alternative in light of its simplicity. The DMFCs utilizing a proton exchange membrane (PEM) have the capability of efficient heat removal and thermal control through the circulating liquid, as well as elimination of the humidification required to prevent membrane dryout. These two characteristics have to be accounted for in both direct and indirect hydrogen systems, as they impact the systems' volume and weight, and consequently the output power density.

Basic Operating Principle

Fig. 3 illustrates the basic operating principle of a DMFC. The DMFC allows for the direct use of an aqueous, low-concentration (3%) liquid methanol solution as the fuel. Air is the oxidant. The methanol and water react directly in the anode chamber of the fuel cell to produce carbon dioxide and protons that permeate the PEM and react with the oxygen at the cathode. The half-cell reactions are

$$\text{Anode}: CH_3OH(l) + H_2O(l) = CO_2(g) + 6H^+ + 6e^-$$

$$\text{Cathode}: 6H^+ + 6e^- + \tfrac{3}{2}O_2(g) = 3H_2O(l)$$

and the net cell reaction is

$$CH_3OH(l) + \tfrac{3}{2}O_2(g) = CO_2(g) + 2H_2O(l).$$

Because the PEM (typically Nafion 117) is used as the electrolyte, the cell operating temperature is usually less than the water boiling temperature to prevent the dryout of the membrane. Typically the operating temperature is

Table 3 Technical progresses made and targets set by Ballard Power Systems for PEM fuel cell stacks operating directly on hydrogen for transportation application

Characteristic	2002	2003	2004	2005	2007	2010
Stack power density (kWe/L)	0.777	0.905	1.205	1.470	1.80	2.50
Cost (US$/kWe)	150	134	103	73	60	30
Start-up to 50% of rated power at the ambient temperature (°C)	−15	−15	−15 (−20)	−25	−25	−30
Cold start-up time (s)	150	50	8 (100)	90	30	30
Durability with cycling (h)	200	700	2100	2100	3200	5000

Note: The numbers shown in the table are evaluated in a manner consistent with those shown in Tables 1 and 2. The numbers for 2002–2005 are the results already achieved; the numbers for 2007 and 2010 are the targets set.
Source: From Ballard Power Systems (see Refs. 2 and 3).

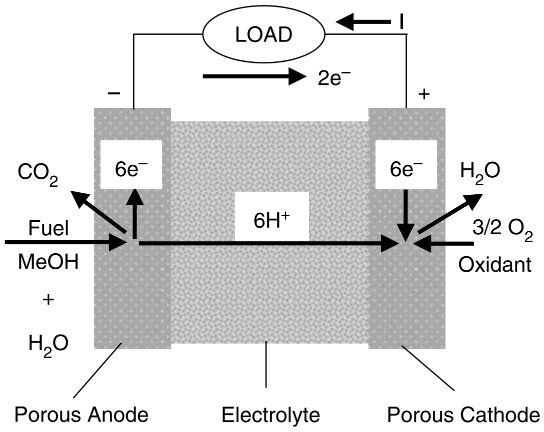


Fig. 3 Schematic of a DFMC.

around 90°C; the operating pressure ranges from one to several atmospheres; and methanol is fed to the cell in liquid form (the so-called liquid-feed DMFCs). Vapor-feed DMFCs have also been considered that operate above 100°C—then heat must be needed to vaporize methanol before feeding into the cell because methanol under the room condition is in liquid form.

Acceptable Contamination Levels

The system is extremely sensitive to CO and hydrogen sulfide (H_2S). Carbon monoxide may exist as one of the reaction intermediaries, and it can poison the catalyst used. There are arguments as to whether CO is present in the anode during the reaction. Sulfur may be present if methanol is made of petroleum oils; if so, the sulfur must be removed. Further, several other reaction intermediaries can poison the catalyst as well.

Major Technological Problems

The PEM used in the DMFCs is Nafion 117, which is the same as employed in the PEMFCs. Although it works well in both types of cells, it is expensive and has only one supplier. Because the electrolyte in DMFCs is essentially acid, expensive precious metals (typically platinum or its alloys) are used as the catalyst. However, the most serious problem is the so-called methanol crossover. This phenomenon is caused by the electro-osmotic effect. When the protons migrate through the electrolyte membrane, a number of water molecules are dragged along with each proton; because methanol is dissolved in liquid water on the anode side, the methanol is dragged through the membrane electrolyte (along with the protons and water) to reach the cathode side. Fortunately, the methanol at the cathode is oxidized into carbon dioxide and water at the cathode catalyst sites, producing no safety hazards. But the methanol oxidation at the cathode does not produce useful electric energy. The development of a new membrane with low methanol crossover is a key to the success of DMFCs.

Such a low methanol crossover membrane has a number of advantages. First, it reduces the methanol crossover, enhancing fuel utilization and increasing energy efficiency. Second, it reduces the amount of water produced at the cathode, leading to a lower activation and concentration polarization, and allowing for a higher cell voltage at the same operating current. Third, it allows for higher methanol concentration in the fuel stream, resulting in better performance.

Technological Status

Direct methanol fuel cells are the least developed of the fuel cell technologies. Though methanol itself has simpler storage requirements than hydrogen and is easier to make and transport, its electrochemical activity is much slower than that of hydrogen; its oxidation rate is about four orders of magnitude smaller than that of hydrogen. Also, the conversion takes place at low temperatures (about 80°C–90°C), and contamination is a serious problem.

The state-of-the-art performance boasts an energy conversion efficiency of 34% (LHV) from methanol to electricity at 90°C using 20 psig air—that is, a cell voltage of 0.5 V (corresponding to a voltage efficiency of 42%) together with the methanol crossover. This accounts for 20% of the current produced (equivalent to a fuel efficiency of 80%). This 20% methanol crossover occurs when the fuel stream used is an aqueous solution containing only 3% methanol. It has been projected that with better membranes being developed by the University of Southern California (USC), as well as improvements in membrane electrode assembly, a cell voltage of 0.6 V can be achieved with only 5% methanol crossover. This is equivalent to 50% voltage efficiency and 95% fuel efficiency, resulting in an overall stack efficiency of 47% (LHV) from methanol to electricity. The DMFC system efficiency will be lower due to the running of auxiliary systems.

The DMFC power system is underdeveloped, and until now, no one could demonstrate any feasibility for commercialization. It remains in the laboratory scale of small demonstrators in the sub-kW range. As such, no system cost estimate is available or has ever been attempted. However, the current DMFCs use essentially the same cell components, materials, construction, and fabrication techniques as the PEMFCs; therefore, it is expected that the system and component costs will be similar to those of the PEMFCs. One company was recently formed to explore the potential of DMFC systems and to develop DMFC for transportation applications. However, judging by the progress of other fuel cells, such systems are at least ten years away from demonstrating any realistic practical applications.

Table 4 Operational characteristics and technological status of various fuel cells

Type of fuel cells	Operating temperature (°C)	Power density (mW/cm^2) (present) projected	Projected rated power level (kW)	Fuel efficiency	Lifetime projected (h)	Capital cost projected (U.S.$/kW)	Application areas
AFC	60–90	(100–200) > 300	10–100	40–60	> 10,000	> 200	Space, transportation
PEMFC	50–80	(350) > 600	1–1000	45–60	5,000	30	Transportation
					> 40,000	> 500	Stationary
					> 2,500	> 1000	Space
DMFC	90	(230) ?	1–100	34	> 10,000	> 200	Transportation

Applications

Direct methanol fuel cells offer a potential for high power density and cold-start capabilities, a convenience for onboard fuel storage, and compatibility with existing refueling infrastructure. Therefore, DMFCs are the most attractive for applications in which storage or generation of hydrogen requires significant effort and has a negative impact on the volume and weight of the system. As a result, DMFCs have great potential for portable and transportation applications, including small appliances, laptop computers, cell phones, automobiles, trains, and ships. In terms of utility applications, small DMFC units will have potential for use in residential and office buildings, hotels, hospitals, etc. Because methanol can be made from agricultural products, its use is also compatible with renewable energy sources to allow for sustainable development.

CONCLUDING REMARKS

A summary of the preceding descriptions of the three major types of low-temperature fuel cells—the AFCs, PEMFCs, and DFMCs—is provided in Table 4, including the operational characteristics and technological status.

For stationary cogeneration, the most economic fuels currently are fossil fuels. Although AFCs demonstrate the best performance when operating on pure hydrogen and oxygen, their intolerance of carbon dioxide eliminates their potential role for utility applications. Proton exchange membrane fuel cells will require complex fuel processing and conditioning subsystems for the reforming of hydrocarbon fuels (notably natural gas), the removal of impurities (including carbon monoxide and sulfur compounds), and external humidification devices. PEMFCs operating directly on hydrogen are believed to be the most promising candidates for transportation applications due to their high power density, fast startup, high efficiency, and easy and safe handling. The main challenges to their commercialization are the cost and durability (or lifetime). PEMFCs will be best for light-duty vehicle applications. Due to the difficulty of onboard fuel (hydrogen) storage and the lack of infrastructure for fuel (hydrogen) distribution, DFMCs are believed by some to be the most appropriate technology for vehicular application, and they are also suited for portable applications due to the high energy density of methanol. PEMFCs are believed to be between five and ten years away from commercialization, while DFMCs are still early in their technological development. Finally, AFCs may stand the best chance for transportation use as far as weight and startup are concerned. Their use is limited only because of their inability to tolerate carbon dioxide contamination in both the fuel and oxidant streams.

REFERENCES

1. *Hydrogen, Fuel Cells and Infrastructure Technologies Program: Multi-Year Research, Development and Demonstration Plan*; U.S. Department of Energy: February 2005.
2. Stone, C. Technology leadership: A roadmap to commercially viable PEMFC stack technology. *Invited Keynote Lecture*, The First International Green Energy Conference, Waterloo, ON, Canada, June 12–16, 2005.
3. Ballard Power Systems. Web site: http://www.ballard.com/be_informed/fuel_cell_technology/roadmap# (accessed on August 8, 2006).

BIBLIOGRAPHY

1. Blomen, L.J.M.J.; Mugerwa, M.N. *Fuel Cell Systems*; Plenum Press: New York, NY, 1993.
2. Larminie, J.; Dicks, A. *Fuel Cell Systems Explained*, 2nd Ed.; Wiley: U.K, 2003.
3. Li, X. *Principles of Fuel Cells*; Taylor & Francis: New York, NY, 2006.
4. Kordesch, K.; Simader, G. *Fuel Cells and Their Applications*; VCH Press: New York, 1996.
5. Appleby, A.J.; Foulkes, F.R. *Fuel Cell Handbook*; Van Nostrand Reinhold Company: New York, NY, 1989.
6. Hoogers, G. *Fuel Cell Technology Handbook*; CRC Press: Boca Raton, FL, 2002.
7. Vielstich, W.; Gasteiger, H.; Lamm, A. *Handbook of Fuel Cells: Fundamentals*; Wiley: New York, NY, 2003; [4-volume set].

Geothermal Energy Resources

Ibrahim Dincer
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology (UOIT), Oshawa, Ontario, Canada

Arif Hepbasli
Department of Mechanical Engineering, Faculty of Engineering, Ege University, Izmir, Turkey

Leyla Ozgener
Department of Mechanical Engineering, Celal Bayar University, Manisa, Turkey

Abstract
This article presents some historical background on geothermal energy resources and their applications, and discusses some technical, energetic, exergetic, economical, environmental and sustainability aspects of geothermal energy systems and their applications. A case study is also presented to compare energy and exergy efficiencies of a geothermal energy system.

NOMENCLATURE

$\dot{E}$	energy rate (kW)
$\dot{E}x$	exergy rate (kW)
$\dot{F}$	exergy rate of the fuel (kW)
h	specific enthalpy (kJ kg^{-1})
$\dot{I}P$	improvement potential rate (kW)
$\dot{m}$	mass flow rate (kg s^{-1})
$\dot{P}$	product exergy rate (kW)
$\dot{Q}$	heat transfer rate (kW)
s	specific entropy (kJ kg^{-1} K^{-1})
SExI	specific exergy index (dimensionless)
T	temperature (°C or K)
$\dot{W}$	work rate, power (kW)

Greek letters

β	exergetic factor (%)
η	energy efficiency (%)
ψ	exergy efficiency (%)
δ	fuel depletion rate (%)
ξ	productivity lack (%)
χ	relative irreversibility (%)

Subscripts

0	reference environment, dead state
d	natural direct discharge
HE	heat exchanger
in	inlet
s	source
out	outlet
tot	total

Keywords: Geothermal; Energy; Exergy; Environment; Sustainability; Performance; System.

INTRODUCTION

The word "geothermal" comes from the Greek words geo (earth) and therme (heat). So, geothermal means earth heat. Geothermal energy is the thermal energy within the earth's crust—the thermal energy in rock and fluid (water, steam, or water containing large amounts of dissolved solids) that fills the pores and fractures within the rock, sand, and gravel. Calculations show that the earth, originating from a completely molten state, would have cooled and become completely solid many thousands of years ago without an energy input in addition to that of the sun.[1] It is believed that the ultimate source of geothermal energy is radioactive decay within the earth.

Geothermal energy utilization is basically divided into two categories, i.e., electric energy production and direct uses. The distribution of the various categories of direct use worldwide is approximately 32% for geothermal heat pumps, 30% for bathing and swimming (including balneology), 20% for space heating (of which 83% is for district heating), 7.5% for greenhouse and open-ground heating, 4% for industrial process heat, 4% for aquaculture pond and raceway heating, <1% for agricultural drying, <1% for snow melting and cooling, and <0.5% for other uses. The equivalent annual savings in fuel oil amounts to 170 million barrels (25.4 million tonnes) and 24 million tonnes in carbon emissions to the atmosphere.[2]

Most of the world's geothermal power plants were essentially built in the 1970s and 1980s as a solution to the 1973 oil crisis. There was an immediate need to generate electricity from alternative energy resources and geothermal energy was essentially given particular attention in this regard.

Lund et al.[2] have recently reported that the world total installed capacity by the end of 2004 for direct use of geothermal energy was 27,825 MW, almost a twofold increase over the 2000 data, growing at a compound rate of

12.9% annually. The total annual geothermal energy utilization became 261,418 TJ (72,622 GWh), almost a 40% increase over the 2000 data, growing at a compound rate of 6.5% annually. During the past ten years, the installed capacity has increased by 12.4%/year and the use has increased by 8.8%/year. The countries with the largest installed capacity and annual energy use are the United States, Sweden, China, Iceland, and Turkey, accounting for about 66% of the installed capacity and 60% of the annual energy use.

In this article, some historical background on geothermal and geothermal energy applications are presented, some technical, economical, environmental and sustainability aspects of geothermal energy are discussed, and performance evaluations tools in terms of energy and exergy analyses for such geothermal energy systems are outlined. A case study is also presented to highlight the importance of exergy use as a potential tool for system analysis, design, and improvement.

HISTORICAL BACKGROUND

Archaeological evidence shows that the first human use of geothermal resources in North America occurred more than 10,000 years ago with the settlement of Paleo-Indians at hot springs.[3] The springs served as a source of warmth and cleansing, with their minerals as a source of healing. While people still soak in shallow pools heated by the earth, engineers are developing technologies that will allow people to probe more than 15 km below the earth's surface in search of geothermal energy.

Here, Table 1 presents a summary table of historical events that have shaped the growth of geothermal energy from its origins to the present.

TECHNICAL ASPECTS OF GEOTHERMAL ENERGY

The details on the technical and fundamental aspects of geothermal energy are well outlined in,[4] and a brief summary of them is presented here.

The discovery that the temperature in deep mines exceeded the surface temperature implied the existence of a source of deep geothermal energy. The average geothermal gradient (or the increase in temperature with depth) near the surface, say within a few km, is about 0.03°C/m, i.e., 30°C/km, but values as low as about 10°C/km are found in ancient continental crust and very high values (>100°C/km) are found in areas of active volcanism.[7]

In some areas, geothermal energy is a viable economic alternative to conventional energy generation. Commercially viable geothermal fields have the same basic structure. The source of heat is generally a magmatic intrusion into Earth's crust. The magma intrusion generally measures 600°C–900°C at a depth of 7–15 km. The bedrock containing the intrusion conducts heat to overlying aquifers (i.e., layers of porous rock such as sandstone that contain significant amounts of water) covered by a dome-shaped layer of impermeable rock such as shale or by an over-lying fault thrust that contains the heated water or steam. A productive geothermal source generally produces about 20 tons (18.1 metric tons) of steam or several hundred tons of hot water per hour. Historically, some heavily exploited geothermal fields have had decreasing yields due to a lack of replenishing water in the aquifer rather than to cooling of the bedrock.

There are three general types of geothermal fields: hot water, wet steam, and dry steam. Hot water fields contain reservoirs of water with temperatures between 60 and 100°C and are most suitable for space heating and agricultural applications. For hot water fields to be commercially viable, they must contain a large amount of water with a temperature of at least 60°C and lie within 2000 m of the surface. Wet steam fields contain water under pressure and usually measure 100°C. These are the most common commercially exploitable fields. When the water is brought to the surface, some of the water flashes into steam and the steam may drive turbines that can produce electrical power. Dry steam fields are geologically similar to wet steam fields, except that superheated steam is extracted from the aquifer. Dry steam fields are relatively uncommon. Because superheated water explosively transforms into steam when exposed to the atmosphere, it is much safer and generally more economical to use geothermal energy to generate electricity, which is much more easily transported.

The thermal efficiency of steam plants is defined as the ratio of the net power generated to the energy of the geothermal fluid at the plant site, and it ranges from 10 to 17%, respectively. These low percentages are due to geothermal resources being at relatively low temperatures. The thermal efficiency of binary cycle plants, defined as the ratio of the net power generated to the energy of the brine at the plant site, ranges from 2.8 to 5.5%, respectively. These percentages are even lower because binary design plants use lower-temperature geothermal resources available for power generation.[8]

To be commercially viable, geothermal electrical generation plants must be located near a large source of easily accessible geothermal energy. A further complication in the practical utilization of geothermal energy is derived from the corrosive properties of most groundwater and steam. In fact, prior to 1950, metallurgy was not advanced enough to enable the manufacture of steam turbine blades resistant to corrosion. Geothermal energy sources for space heating and agriculture have been used extensively in Iceland, and to some degree, Japan, New Zealand, and the former Soviet Union countries. Other applications include paper manufacturing and water desalination.

Table 1 Some worldwide historical events of geothermal energy, including some recent activities in Canada

Date	Event
Prehistory	Use of caves' geothermal energy for winter warmth and summer cooling; people gathering at hot springs to enjoy warmth from deeper geothermal resources
Early history	The Romans used geothermally heated water in their bathhouses for centuries. The Romans also used the water to treat illnesses and heat homes. In Iceland and New Zealand, many people cooked their food using geothermal heat. Some North American native tribes also used geothermal vents for both comfort heat and cooking temperatures. Most of these early uses of the earth's heat were through the exploitation of geothermal vents. Early peoples also utilize near-constant subsurface temperatures (geothermal energy) to keep vegetables in root cellars from freezing or spoiling
1860s	Energy from hot springs begins to be used in the United States to heat homes and bathhouses
1886	Hot water from the Banff, Alberta hot springs piped to hotels and spas
1904	Italian scientist Piero Ginori Conti invents the first geothermal electric power plant
1946	First ground-source geothermal heat pump installed at the Commonwealth Building in Portland, Oregon
1958	The first turbine-generator sets in Wairakei "A" Station were synchronized to the national grid in New Zealand—the first geothermal electrical development in the world using "wet steam"
1960	Pacific Gas and Electric begins operation of first large-scale geothermal power plant at the Geysers north of San Francisco, producing 11 megawatts. 69 geothermal generating facilities are in operation at 18 resource sites around the United States
1970s	During the oil crisis of 1973, many countries begin exploring alternative sources of energy, including geothermal energy. This sparks government-sponsored renewable energy research programs in Germany, Sweden, Canada, the United Kingdom, and the United States. Geothermal heat pumps and underground thermal energy storage systems gain popularity as a means to reduce heating and cooling costs, especially in commercial and institutional buildings
1975	Drilling begins to assess high-temperature geothermal resources for electricity generation in British Columbia
1976–1986	Ten-year federal research program assesses geothermal energy resources, technologies, and opportunities for Canada. In the early 1980s, some plants in Indonesia, Kenya, Turkey, the Philippines, and Portugal (Azores) were built
1990s	As public concerns about environmental issues such as air pollution and climate change grow, governments in Canada and elsewhere take a greater interest in using renewable energy as a way to decrease greenhouse gases and other emissions
1990	Ontario Hydro funds a program to install geothermal heat pumps in 6749 residences not served by natural gas
2004	Western GeoPower Corp. applies for government approvals to build a $340 million, 100-megawatt geothermal power plant at Meager Creek, northwest of Whistler, B.C., which could begin producing power as early as 2007. Manitoba government announces program to provide loans of to $15,000 towards installation of geothermal heat pump systems
2005	There are 25 countries producing over 54,000 GWh/year of electricity from geothermal energy from over 8700 MW_e of installed capacity

Source: From Refs. 5 and 6.

In summary, while geothermal energy is generally considered a nonpolluting energy source, water from geothermal fields may often contain some amounts of hydrogen sulfide and dissolved metals, making its disposal difficult. That is why depending on the location and area, a careful treatment is required.

Here, Table 2 presents a summary of the types of geothermal resources, reservoirs temperatures, reservoirs fluids, applications, and systems.

GEOTHERMAL ENERGY AND ENVIRONMENTAL IMPACT

Problems with energy supply and use are related not only to global warming but also to such environmental concerns as air pollution, acid precipitation, ozone depletion, forest destruction, the emission of radioactive substances, etc. These issues must be taken into consideration simultaneously if humanity is to achieve a bright energy future with minimal environmental impacts. Much evidence exists that suggests the future will be negatively impacted if humans keep degrading the environment.[9]

One solution to both energy and environmental problems is to make better use of renewable energy resources, particularly geothermal resources. This cause is sometimes espoused with a fervor which leads to extravagant and impossible claims being made. Engineering practicality, reliability, applicability, economy, scarcity of supply, and public acceptability should all be considered accordingly.[10]

In order to achieve the energetic, economic, and environmental benefits that geothermal resources offer, the following integrated set of activities should be acted accordingly:

1. *Research and Development*: Research and development priorities should be set in close consultation with industry to reflect their needs. Most research is

Table 2 Geothermal resource types, applications and system types

Reservoir temperature	Reservoir fluid	Applications	Systems
High temperature (>220°C)	Water or steam	Power generation	Flash steam
			Combined (flash and binary) cycle
		Direct use	Direct fluid use
			Heat exchangers
			Heat pumps
Medium temperature (100°C–220°C)	Water	Power generation	Binary cycle
		Direct use	Direct fluid use
			Heat exchangers
			Heat pumps
Low temperature (50°C–150°C)	Water	Direct use	Direct fluid use
			Heat exchangers
			Heat pumps

Source: Adapted from World Bank (see Ref. 11).

conducted through cost-shared agreements and falls within the short-to-medium term. Partners in these activities should include a variety of stakeholders in the energy industry, such as private sector firms, utilities across the country, provincial governments, and other federal departments.

2. *Technology Assessment*: Appropriate technical data should be gathered in the lab and through field trials on factors such as cost benefit, reliability, environmental impact, safety, and opportunities for improvement. These data should also assist the preparation of technology status overviews and strategic plans for further research and development.
3. *Standards Development*: The development of technical and safety standards is needed to encourage the acceptance of proven technologies in the marketplace. Standards development should be conducted in cooperation with national and international standards-writing organizations as well as other national and provincial regulatory bodies.
4. *Technology Transfer*: Research and development results should be transferred through the sponsorship of technical workshops, seminars and conferences, as well as through the development of training manuals and design tools, web tools, and the publication of technical reports.

Such activities will also encourage potential users to consider the benefits of adopting geothermal energy technologies. In support of developing near-term markets, a key technology transfer area is to accelerate the use of geothermal energy technologies, particularly for district heating applications and power generation.

GEOTHERMAL ENERGY AND SUSTAINABILITY

Sustainable development requires a sustainable supply of clean and affordable energy resources that do not cause negative societal impacts.[12] Supplies of such energy resources as fossil fuels and uranium are finite. Energy resources such as geothermal resources are generally considered renewable and therefore sustainable over the relatively long term.

Sustainability often leads local and national authorities to incorporate environmental considerations into energy planning. The need to satisfy basic human needs and aspirations combined with increasing world population will make the need for successful implementation of sustainable development increasingly apparent. Various criteria that are essential to achieving sustainable development in a society include:

- Information about and public awareness of the benefits of sustainability investments,
- Environmental education and training,
- Appropriate energy and exergy strategies,
- The availability of renewable energy sources and cleaner technologies,
- A reasonable supply of financing, and
- Monitoring and evaluation tools.

Geothermal as a renewable energy has an important role to play in meeting future energy needs in both rural and urban areas. The development and utilization of geothermal energy applications should be given a high priority, especially in light of increased awareness of the adverse environmental impacts of fossil-based generation. The need for sustainable energy development is increasing rapidly in the world. Widespread use of geothermal energy is important for achieving sustainability in the energy sectors in both developing and industrialized countries.

Geothermal energy can be a key component for sustainable development for three main reasons, as follows:

- They may generally cause environmental impact much less than other energy resources. The variety of geothermal energy applications provides a flexible array of options for their use.
- They are considered renewable, and such energy resources can provide a reliable and sustainable supply of energy almost indefinitely. In contrast, fossil fuel and uranium resources are diminished by extraction and consumption.
- They favor both system centralization and decentralization and local solutions that are somewhat independent of the national network, thus enhancing the flexibility of the system and providing economic benefits to small isolated populations. Also, the small scale of the equipment, depending on the application, often reduces the time required from initial design to operation, providing greater adaptability in responding to unpredictable growth and changes in energy demand.

Not all renewable energy resources are inherently clean, meaning they cause no burden on the environment in terms of waste emissions, resource extraction, or other environmental disruptions. Nevertheless, use of geothermal resources almost certainly can provide a cleaner and more sustainable energy system than increased controls on conventional energy systems.

To seize the opportunities, it is essential to establish a geothermal energy market and gradually build up the experience with the cutting-edge technologies. The barriers and constraints to the diffusion of geothermal energy use should be removed. The legal, administrative, and financing infrastructure should be established to facilitate planning and the application of geothermal energy projects. Government can and should play a useful role in promoting geothermal energy technologies through funding and incentives to encourage research and development as well as commercialization and implementation in both urban and rural areas.

PERFORMANCE EVALUATION

From the thermodynamics point of view, exergy is defined as the maximum amount of work that can be produced by a system or a flow of matter or energy as it comes to equilibrium with a reference environment. Exergy is a measure of the potential of the system or flow to cause change as a consequence of not being completely in stable equilibrium relative to the reference environment. Unlike energy, exergy is not subject to a conservation law (except for ideal, or reversible, processes). Rather, exergy is consumed or destroyed due to irreversibilities in any real process. The exergy consumption during a process is proportional to the entropy created due to irreversibilities associated with the process.

Exergy analysis is a technique that uses the conservation of mass and conservation of energy principles together with the second law of thermodynamics for the analysis, design, and improvement of geothermal energy systems as well as other systems. It is also useful for improving the efficiency of energy-resource use, for it quantifies the locations, types, and magnitudes of wastes and losses. In general, more meaningful efficiencies are evaluated with exergy analysis rather than energy analysis because exergy efficiencies are always a measure of how nearly the efficiency of a process approaches the ideal. Therefore, exergy analysis identifies accurately the margin available to design more efficient energy systems by reducing inefficiencies. We can suggest that thermodynamic performance is best evaluated using exergy analysis because it provides more insights and is more useful in efficiency-improvement efforts than energy analysis. For exergy analysis, the characteristics of a reference environment must be specified. This is commonly done by specifying the temperature, pressure, and chemical composition of the reference environment. The results of exergy analyses, consequently, are relative to the specified reference environment, which in most applications is modeled after the actual local environment. The exergy of a system is zero when it is in equilibrium with the reference environment.

Energy and exergy balances for an unsteady-flow process in a system during a finite time interval can be written as:

$$\text{Energy input} - \text{Energy output} = \text{Energy accumulation} \quad (1)$$

$$\text{Exergy input} - \text{Exergy output} - \text{Exergy consumption} = \text{Exergy accumulation} \quad (2)$$

For a general steady-state, steady-flow process, it is assumed that the accumulation terms in the above equations become zero. Therefore, input terms become equal to output terms.

For a steady-state, steady-flow process, the mass balance equation can be expressed as

$$\Sigma \dot{m}_{in} = \Sigma \dot{m}_{out} \quad (3)$$

For a general steady-state, steady-flow process, the general energy and exergy balance equations can also be written more explicitly as

$$\dot{Q} + \Sigma \dot{m}_{in} h_{in} = \dot{W} + \Sigma \dot{m}_{out} h_{out} \quad (4)$$

and

$$\Sigma \left(1 - \frac{T_0}{T_s}\right) \dot{Q}_s + \Sigma \dot{m}_{in}(h_{in} - T_0 s_{in}) + \dot{E}x_d = \dot{W} + \Sigma \dot{m}_{out}(h_{out} - T_0 s_{out}) \quad (5)$$

We can generally define the energy and exergy efficiencies as follows

$$\eta = \frac{\dot{E}_{net}}{\dot{E}_{tot}} \quad (6)$$

and:

$$\psi = \frac{\dot{E}x_{net}}{\dot{E}x_{tot}} \quad (7)$$

For further details of the energy and exergy analysis, see Dincer and Rosen.[12]

In addition to the above analysis, several researchers[13–17] have introduced some resource classification criteria and performance and quality parameters as explained below:

Geothermal resources are classified as low, intermediate, and high enthalpy resources according to their reservoir temperatures, as mentioned earlier. The ranges may vary from one researcher to another. For example, Dickson and Fanelli[13] classified them as low (<90°C), medium (between 90 and 150°C) and high (>150°C), respectively. However, Lee et al. (2001) have come up with another approach based on their quality rather than their temperature through specific exergy index (SExI), as follows:

$$\text{SExI} = \frac{h_{brine} - 273.16 s_{brine}}{1192} \quad (8)$$

with the following criteria:

- Specific exergy index < 0.05 for low quality geothermal resources
- 0.05 ≤ SExI < 0.5 for medium quality geothermal resources
- Specific exergy index ≥ 0.5 for high quality geothermal resources

(Note that here, the demarcation limits for these indices are exergies of saturated water and dry saturated steam at 1 bar absolute.)

van Gool[18] has also noted that maximum improvement in the exergy efficiency for a process or system is obviously achieved when the exergy loss or irreversibility ($\dot{E}x_{in} - \dot{E}x_{out}$) is minimized. Consequently, he suggested the concept of an exergetic "improvement potential" when analyzing different processes. This improvement potential on the rate basis, denoted $\dot{\text{I}}\text{P}$, is given by[16]:

$$\dot{\text{I}}\text{P} = (1 - \psi)(\dot{E}x_{in} - \dot{E}x_{out}) \quad (9)$$

Xiang et al.[17] have also introduced some additional thermodynamic parameters to use for geothermal energy applications, as follows:

- Fuel depletion rate : $\delta_i = \frac{\dot{E}x_i}{\dot{F}_{tot}}$ (10)
- Relative irreversibility : $\chi_i = \frac{\dot{E}x_i}{\dot{E}x_{tot}}$ (11)
- Productivity factor : $\xi_i = \frac{\dot{E}x_i}{\dot{P}_{tot}}$ (12)
- Exergetic factor : $\beta_i = \frac{\dot{F}_i}{\dot{F}_{tot}}$ (13)

Note that in the calculations, the temperature T_0 and pressure P_0 of the environment are often taken as

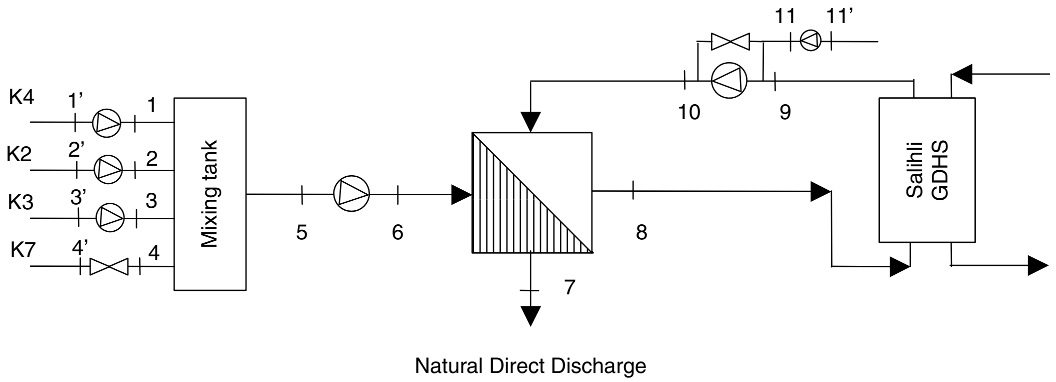


Fig. 1 A schematic of the Salihli geothermal district heating system (SGDHS).

standard-state values, such as 1 atm and 25°C. However, these properties may be specified differently depending on the application. T_0 and P_0 might be taken as the average ambient temperature and pressure, respectively, for the location at which the system under consideration operates. Or, if the system uses atmospheric air, T_0 might be specified as the average air temperature. If both air and water from the natural surroundings are used, T_0 would be specified as the lower of the average temperatures for air and water.

Detailed energy and exergy as well as exergoeconomic analysis methodologies for geothermal energy systems are given elsewhere.[19–22]

CASE STUDY

Here, a case study analysis of Salihli geothermal district heating system (SGDHS), which has a maximal yield of 87 L/s at an average reservoir temperature of 95°C with a minimal capacity of 838 MW[19] is presented. The SGDHS was originally designed for 20,000 residences equivalence. Of these, 2400 residences equivalence are heated by geothermal energy as of February 2004. The outdoor and indoor design temperatures for the system are 4 and 22°C, respectively.

Fig. 1 illustrates a schematic of the SGDHS where two hospitals and official buildings heated by geothermal energy were also included. The SGDHS consists mainly of three cycles, such as: (a) the energy production cycle (geothermal well loop and geothermal heating center loop), (b) the energy distribution cycle (district heating distribution network), and (c) the energy consumption cycle (building substations). At the beginning of 2004, there were seven wells ranging in depth from 70 to 262 m in the SGDHS. Of these, five wells were in operation at the date studied and two wells (K5 and K6) were out of operation. Four wells (designated as K2–K4, and K7) and one well (K1) are production and balneology wells, respectively.

The well head temperatures of the production wells vary from 56 to 115°C while the volumetric flow rates of the wells range from 2 to 20 L/s. The geothermal fluid is basically sent to two primary plate type heat exchangers and is cooled to about 45°C as its heat is transferred to secondary fluid.

The geothermal fluid (7) is discharged via natural direct discharge with no recharge to the Salihli geothermal field production, but reinjection studies are expected to be completed in the near future. The temperatures obtained during the operation of the SGDHS are, on average, 98/45°C for the district heating distribution network and 62/42°C for the building circuit. By using the control valves for flow rate and temperature at the building main station, the needed amount of water is sent to each housing unit and the heat balance of the system is achieved.

Table 3 Salihli geothermal district heating system (SGDHS) data and energetic, exergetic, and thermodynamics parameter calculated

Item No	Component	Exergy destruction rate (kW)	Utilized power (kW)	Installed capacity (kW)	$\dot{P}$ (kW)	$\dot{F}$ (kW)	Relative irreversibility χ (%)	Fuel depletion rate δ (%)	Productivity factor ξ (%)	Exergetic factor β (%)	Energy efficiency η (%)	Exergy efficiency ψ (%)
1	Heat exchanger	458.46	10,226.83	43,961.38	1524	1982.36	44.09	20.69	30.08	89.49	—	76.87
2	K4 well pump	21.46	24.75	55	3.29	24.75	2.06	0.96	1.40	1.12	65–80	13.29
3	K2 well pump	12.74	20.25	45	7.51	20.25	1.22	0.57	0.83	0.92	65–80	37.09
4	K3 well pump	5.23	20.25	45	15.02	20.25	0.50	0.23	0.34	0.92	65–80	74.17
5	K7 well pump	—	—	—	—	—	—	—	—	—	—	—
6	Salihli booster pump	7.3	55	675	47.7	55	0.70	0.32	0.47	2.48	65–80	86.72
7	Salihli circulation pump	10.28	112.5	537	102.22	112.5	0.98	0.46	0.67	5.07	65–80	90.86
8	Heat exchangers and pumps	515.47	10,468.52	45,520.58	1524	1982.36	—	—	—	—	—	—
9	Overall system[a]	1039.67	1,524	45,520.58	1524	1982.36					55.5	59.44

[a]Based on the exergy (or energy) input to thermal water and water.

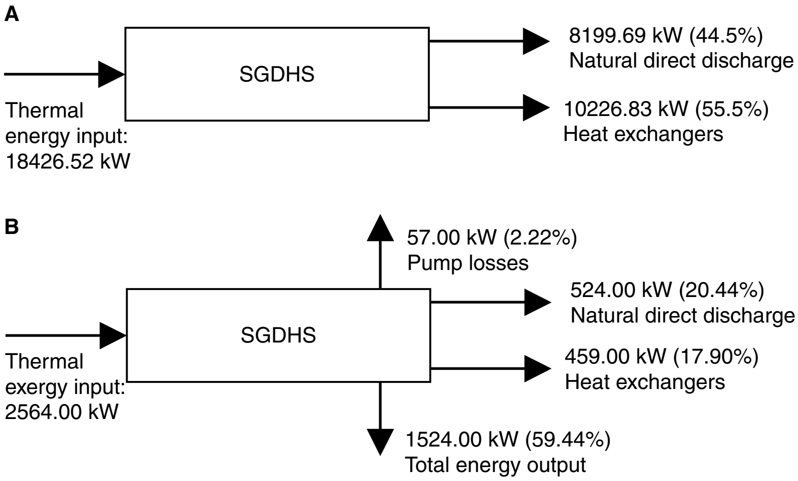


Fig. 2 System flow diagrams for (A) energy and (B) exergy.

Geothermal fluid collected from the four production wells at an average well heat temperature of 95.5°C is pumped to the inlet of the heat exchanger mixing tank later a main collector (from four production wells) with a total mass flow rate of about 47.62 kg/s. Geothermal fluid of intermingling molecules of different species through molecular diffusion was neglected in this study. As a result, not only irreversibility of the mixing tank was assumed equal to zero, but also heat losses from the tank and main collector pipe line (5–6) through the mixing process were neglected.

In the SGDHS investigated, the thermal data on pressures, temperatures, and flow rates of the fluids were measured using a number of devices and employed in the models. The detailed information is presented in Ozgener et al.[19]

RESULTS AND DISCUSSION

The thermodynamic data used in the case study are based on the actual data taken from the SGDHS on 1 February 2004. The number of the wells in operation in the Salihli geothermal field may normally vary depending on the heating days as well as the operating strategy. Taking into account the four productive wells when this study was conducted and using Eq. 8, the SExI is found to be 0.049, which is very close to the limit of the medium quality geothermal resources. This represents that the Salihli geothermal field falls into the low quality geothermal resources according to the classification of Lee.[15]

The total mass flow rates of the geothermal fluid at the inlet of the heat exchanger (the total mass flow rate of the production wells) were measured to be 47.62 kg/s at an average temperature of 44.2°C on 1 February 2004. If hot water losses measured in the district heating distribution network are above 5 m^3/h, the water is generally added via a pump (using pressurized water tanks) to the network in order to cover the leaks taking place. However, the hot water losses were neglected in this study as these losses were below 5 m^3/h on the day of the study.

The calculation results, based on the actual data taken, are presented in Table 3. These include all energetic and exergetic parameters as well as other thermodynamic parameters along with the efficiencies. The energy and exergy efficiencies of the SGDHS are determined to be 55.5 and 59.4%, respectively. Here, the exergy efficiency is higher due to the recirculation and heat recovery processes.

The energy and exergy flow diagrams are presented in Fig. 2. As illustrated in Fig. 2a, the natural direct discharge of the system accounts for 44.5% of the total energy input. An investigation of the exergy flow diagram given in Fig. 2b shows that 40.56% (accounting for about 1040 kW) of the exergy entering the system is lost, while the remaining 59.44% is utilized.

The highest exergy loss of 20.44% occurs from the natural direct discharge in this study. The second largest exergy destruction occurs from the heat exchanger with 17.90% (about 459 kW) of the total exergy input. This is followed by the total exergy destruction associated with the pumps amounting to some 57 kW, which accounts for 2.22% of the exergy input to the system.

In addition, the exergetic improvement potential (IP) rate of Van Gool's[18] is found to be 106.04 kW for the plate-type heat exchanger installed in the SGDHS. Furthermore, in order to improve the system efficiency, natural direct discharge should be decreased.

CONCLUSIONS

In this article, some historical background on geothermal energy resources and applications have been presented, some technical, economical, environmental, and sustainability aspects of geothermal energy have been discussed, and performance evaluations tools in terms of energy and exergy analyses for such geothermal energy systems have been outlined. A case study has also been presented to highlight the importance of exergy use as a potential tool for system analysis, design, and improvement.

REFERENCES

1. Bullard, E. *Basic Theories, (Geothermal Energy; Review of Research and Development)*; UNESCO: Paris, 1973.
2. Lund, J.W. Freeston, D.H.; Boyd, T.L. World-wide direct uses of geothermal energy 2005. *Proceedings of the World Geothermal Congress 2005*, April 24–29, 2005, Antalya, Turkey, 2005.
3. DOE. Geothermal Technologies Program. U.S. Department of Energy, from the web http://www.eere.energy.gov/geothermal/history.html (Retrieved October 23, 2005).
4. Center for Energy. Geothermal Energy Timeline. from the web http://www.centreforenergy.com/silos/geothermal/geothermalEnvironment/geothermalEnvHistory.asp (Retrieved October 23, 2005).
5. Lund, J.W.. 100 years of geothermal power product. Proceedings Thirtieth Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, California, January 31–February 2, 2005, SGP-TR-176, available at: http://iga.igg.cnr.it/index.php.
6. BookRags. Geothermal Energy. BookRags, from the World Wide web. http://www.bookrags.com/sciences/earthscience/geothermal-energy-woes-01.html (Retrieved October 23, 2005).
7. Barbier, E. Nature and technology of geothermal energy: A review. Renew. Sust. Energ. Rev. **1997**, *1* (1/2), 1–69.
8. Kanoglu, M.; Cengel, Y.A. Economic evaluation of geothermal power generation, heating, and cooling. Energy **1999**, *24*, 501–509.
9. Dincer, I. On energy conservation policies and implementation practices. Int J. Energy Res. **2003**, *27* (7), 687–702.
10. Dincer, I. Renewable energy and sustainable development: a crucial review. Renew. Sust. Energ. Rev. **2000**, *4* (2), 157–175.
11. World Bank. Geothermal Energy, from the Web http://www.worldbank.org/html/fpd/energy/geothermal/technology.htm (Retrieved 23 October 2005).
12. Dincer, I.; Rosen, M.A. Thermodynamic aspects of renewables and sustainable development. Renew. Sust. Energ. Rev. **2005**, *9* (2), 169–189.
13. Dickson, M.H.; Fanelli, M. Geothermal energy and its utilization. In *Small Geothermal Resources*; Dickson, M.H., Fanelli, M., Eds., UNITAR/UNDP Centre for Small Energy Resources: Rome, Italy, 1990; 1–29.
14. Quijano J. Exergy analysis for the Ahuachapan and Berlin Geothermal fields, El Salvador. *Proceedings World Geothermal Congress*, May 28–June 10, 2000, Kyushu-Tohoku, Japan, 2000.
15. Lee, K.C. Classification of geothermal resources by exergy. Geothermics **2001**, *30*, 431–442.
16. Hammond, G.P.; Stapleton, A.J. Exergy analysis of the United Kingdom energy system. Proc. Instn Mech. Engrs. **2001**, *215* (2), 141–162.
17. Xiang, J.Y.; Cali, M.; Santarelli, M. Calculation for physical and chemical exergy of flows in systems elaborating mixed-phase flows and a case study in an IRSOFC plant. Int. J. Energy Res. **2004**, *28*, 101–115.
18. van Gool, W. Energy policy: fairy tales and factualities. In *Innovation and Technology-Strategies and Policies*; Soares, O.D.D.; Martins da Cruz, A.; Costa Pereira, G.; Soares, I.M.R.T.; Reis, A.J.P.S.; Eds.; Kluwer: Dordrecht, 1997; 93–105.
19. Ozgener, L.; Hepbasli, A.; Dincer, I. Energy and exergy analysis of Salihli geothermal district heating system in Manisa, Turkey. Int. J. Energy Res. **2005**, *29* (5), 393–408.
20. Ozgener, L.; Hepbasli, A.; Dincer, I. Energy and exergy analysis of geothermal district heating systems: an application. Build. Environ. **2005**, *40* (10), 1309–1322.
21. Ozgener, L.; Hepbasli, A.; Dincer, I. Energy and exergy analysis of the Gonen geothermal district heating system in Turkey. Geothermics **2005**, *34* (5), 632–645.
22. Ozgener, L.; Hepbasli, A.; Dincer, I. Thermodynamic analysis of a geothermal district heating system. Int. J. Exergy **2005**, *2* (3), 231–245.

Foss–Gre

Geothermal Heat Pump Systems

Greg Tinkler
RLB Consulting Engineers, Houston, Texas, U.S.A.

Foss–Gre

Abstract
Geothermal heat exchanger technology is the most efficient method of heating, cooling, or refrigerating any enclosure that can be conditioned. The earth's thermal mass has an almost-endless capacity to absorb and store heat. Geothermal heat pumps (GHPs) first became well known in the energy crises of the 1970s in the United States; they had been used in Europe for many years before that. Since then, this technology had been researched and implemented successfully in projects of all sizes—from small mobile homes to large schools and universities and to complete developments with multiple systems. These systems can refrigerate down to below 0°F and heat water for domestic use up to 120°F. Geothermal systems have been applied successfully to projects around the world, from inside the Arctic Circle to the desert regions. Geothermal systems have been applied to projects from large buildings in the country to medium-size structures in Manhattan.

The larger the demand placed on an heating, ventilation, air conditioning and refrigeration (HVAC/R) system, the faster the payback period for the upfront capital cost. Vertical geothermal heat pump systems can usually be built for about the same cost as a 4-pipe chiller/boiler system.

Geothermal technologies are considered to be renewable energy as well as energy-efficient technologies. These systems should never be overlooked when thinking about designing any system for the purpose of conditioning space or the temperature of a material or process. These systems are not complicated and can be among the simplest systems for the purposes to which they are applied. This entry explains some of the basic concepts and terms used in this industry.

INTRODUCTION

Use of the earth's resources as a medium for transferring heat is a technology that can be utilized for heating, cooling, and refrigeration. The technology can be referred to as geoexchange systems, ground source heat pumps, ground coupled heat pumps, or earth coupled systems. These interchangeable names all refer to the use of geothermal heat pump (GHP) systems.

The U.S. Department of Energy (Environmental Protection Agency (EPA)) has proclaimed that GHP systems are the most efficient technology available for space conditioning. Typical savings on the total energy use of the building can be from 10 to 50%, depending upon the demand placed on the space conditioning systems. Energy savings realized against the most efficient water-cooled chiller systems can be 10%, whereas savings against typical air-to-air technologies can be as high as 50%. Geothermal heat pump systems in most states have been declared not only energy efficient, but also renewable energy. These distinctions qualify the GHP system for most state energy requirements and/or tax incentives. Geothermal heat pump systems contribute to points necessary for Leadership in Energy and Environmental Design (LEED) certification.

Geothermal heat pump systems are becoming more widely used because they are both renewable energy and energy efficient. This is one technology that fits into both categories. This article is intended to describe the technology as well as the benefits of using these systems.

For a discussion of this topic, some terms and conversion rates should be defined.

Keywords: Geothermal; Geoexchange; Ground source heat pump; Earth heat exchanger; Ground coupled heat pump; Closed loop system; Pond heat exchanger.

TERMS

British thermal unit (Btu) is the amount of heat that it takes to raise 1 lb of water 1°F. This is the typical unit of measure for heating systems in the United States.

Ton (ton) is 12,000 Btu/h and is defined as the amount of heat that it takes per hour to melt 1 ton of ice in a 24-hour period. In the United States, most "air-cooled" cooling systems are sized or measured based on how many tons of heat they can transfer. A ton is a unit of flow and not a unit of volume. Ton-hour is the expression of volumetric heat.

Kilowatt (kW) is a unit of measure for both heat and electricity. With electricity, a kilowatt is 1000 W. A watt is calculated by multiplying volts times amps for air conditioning systems with only resistive loads. With heat, 1 kW = 3412.83 Btu. Kilowatts are a unit of flow and not a unit of volume. Volume is usually expressed in kilowatt per hour. In European and other countries, the term kilowatt is used to express the amount of heat that the cooling system or heating system can move.

Coefficient of performance (COP) is a relationship of watts-out to watts-in or kilowatts-out divided by kilowatts-in. Receiving 4 kW of heat from a machine by buying 1 kW of energy from the electric plant to operate the machine, for example, represents a COP equal to 4. You may see this term used in both the United States and Europe as a unit of measure of the performance of a heating and/or cooling system.

Energy efficiency ratio (EER) is a term used to describe the efficiency of a cooling system in the United States. The EER is computed by dividing the british thermal unit per hour that the unit produces by the amount of watts bought from the utility company to run that machine. When attempting to discover the EER listed on a piece of equipment, a seasonal energy efficient ratio (SEER) may be listed instead. The SEER is not the same as the EER. The SEER is not an easily calculated value because it requires a run-test at some temperature condition, of the air streams, to determine the value. The SEER may not represent the actual performance when installed on a project.

Greenhouse gas (GHG) is the byproduct/pollution generated either from a furnace/boiler or from a power plant burning a fossil fuel to generate the electricity used by the machine. Typical GHG produced are carbon dioxide (CO_2), sulfur dioxide (SO_2), and nitrous oxide (NOx). Cars produce CO_2, for example, which is a source of pollution in most metropolitan areas. Trees consume CO_2 and emit O_2. Greenhouse gas pollution is sometimes expressed in the equivalent number of trees it would take to neutralize the pollutant or in the equivalent number of cars it would take to release the same amount of CO_2.

THE GHP SYSTEM

The principle of a GHP system is to transfer heat to and from the earth. In cool weather, the earth's natural heat is collected through a series of pipes called the loop. A fluid circulating in the loop (usually water) carries this heat from the earth to the GHP. Then, an indoor GHP unit uses electrically driven compressors and heat exchangers to concentrate the earth's heat and release it inside the facility at a higher temperature. In typical systems, ducted fans distribute the heat to various rooms.

In warm weather, the process is reversed to cool the facility. This is done by drawing excess heat from the facility; transferring the heat to the fluid using the GHP; and then expelling the heat to the loop, where it is absorbed by the earth. Geothermal heat pump units provide cooling in the same way that a refrigerator operates—by drawing heat from the interior, not by injecting cold air.

Example System 1

Tinkler and Nilsson[1,2] provide the following example, which describes the operation of a residential GHP system and the related measures of COP, energy usage, and efficiency.

The average COP of a typical GHP system during the first ten years is:

- Heating mode—COP: 4.5
- Cooling mode—COP: 4.

Heating Mode

To produce 4.5 kW (15.3 thousand Btu/h) of heating in a building, the GHP system uses 1 kW of electricity and 3.5 kW (11.9 thousand Btu/h) of the earth's natural energy (Fig. 1).

[4.5 = 1 + 3.5]

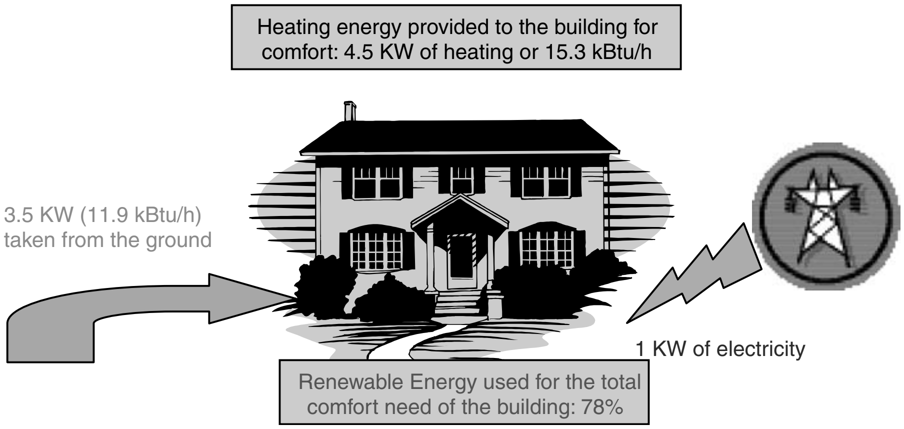


Fig. 1 Operation of a residential Geothermal heat pump (GHP) system: Heating mode.

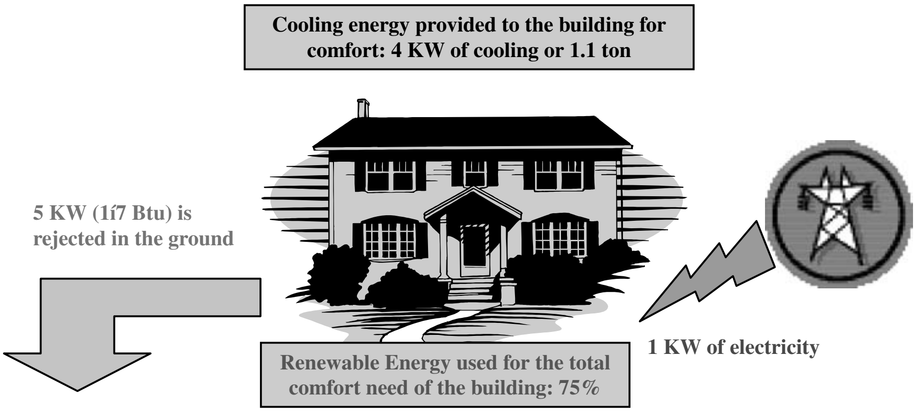


Fig. 2 Operation of a residential Geothermal heat pump (GHP) system: Cooling mode.

The GHP system's COP=4.5 when a conventional gas furnace has a COP of 0.8.

Cooling Mode

To produce 4 kW (1.1 ton) of cooling in a building, the GHP system uses 1 kW of electricity to reject 5 kW (1.7 kBtu/h) in the earth (Fig. 2).

$$[4 = 1 + (-5)]$$

The GHP system has an EER of 14 when a conventional air conditioner (air-cooled) system would have an EER of 9.[1]

Example System 2

An environmental benefit of a GHP system is GHG emission reduction. To illustrate this benefit, consider a 350-ton HVAC retrofit project in Texas, at Lubbock Christian University. The annual emission reduction would be

Carbon dioxide (CO_2) in pounds	Sulfur dioxide (SO_2) in grams	Nitrogen oxide (NO_X) in grams
9,71,998	6,17,039	8,61,661

This is equivalent to preserving 3.61 acres of land from deforestation or taking 95 passenger cars off the road for a year.[2]

These emissions reductions are calculated using System Analyzer software from Trane (a division of American Standard Companies, Inc.). They are based on the electricity and gas savings and on the energy mix of the state. The conversion for cars and trees can be determined by using the EPA Web site.[3]

GHP Compared with Conventional HVAC

The American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. (ASHRAE) reported typical findings for GHP systems compared with conventional HVAC systems:

- *Energy*. Geothermal can reduce energy demand (kW) and energy consumption (kW/h) by 40% compared with air source heat pumps and by more than 70% compared with electric resistance heating with standard air conditioning equipment, according to the EPA. Usually, HVAC accounts for approximately 60%-plus of the annual energy bills of a building.
- *Operation and maintenance*. Geothermal (GHP) can reduce up to six times the costs of the HVAC operation and maintenance costs (Table 1).

Table 1 Findings for GHP systems compared with conventional HVAC system.

	Ground source heat pump (closed loop)	Split system heat pump	Reciprocating chiller + gas-fired boiler	Water source heat pumps + gas-fired boiler
Total maintenance costs per year per cooling ft^2 (cents/year/cooled ft^2)	8.9	34.5	56	63.3
Comparison with ground source heat pump system	—	+288%	+529%	+611%

Source: Study American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. (ASHRAE) 2000 based on a sample of 170 buildings-update on maintenance and services costs of commercial buildings ground source heat pump system (see Ref. 4).

- *System life*. According to ASHRAE, geothermal cooling and heating indoor equipment has a rated equipment life of more than 25 years. The exterior portion of the system—the underground loop field—has a rated materials life of more than 100 years.[4]

TYPES OF EARTH HEAT EXCHANGERS

The term geothermal usually applies to the technology of using water source heat pumps (extended range) connected to a thermal mass. The most widely used form is composed of several vertical boreholes, tied in parallel, that are drilled deep into the earth. The thermal mass of the earth is used as the heat sink/heat source, depending on the climate and the application. Other thermal mass technologies are ponds or lakes with a heat exchanger in the bottom of the body of water. A third form is horizontal trenches up to 6 ft deep with several looped pipes buried in those trenches. A final form is the use of groundwater volumes pumped from below the earth's surface up to a heat exchanger in the building.

There may also be a combination of these forms on any site. On the Lubbock Christian University campus, for example, are three of the four examples shown in Fig. 3: a vertical closed loop, a water supply and reinjection well, and a pond loop (currently under construction).

Over the past few decades, the commercial/industrial/institutional markets have grown significantly. In the commercial market, large metropolitan areas offer the greatest challenge for developing geothermal systems because of the access to land to tap the earth's thermal mass. To combat this challenge, commercial projects have installed groundwater systems using a local surface water source. When this is not feasible, they have turned to more advanced geothermal systems, such as standing column well technology or hybrid systems tied to a cooling tower or boiler to supplement the geothermal heat exchanger.

Another resourceful solution uses the high-rise building's dewatering system. In several areas of the country, high-rise buildings have a dewatering system to keep their underground structure from flooding. This is an excellent water stream that can be a significant volume of water in which to transfer heat.

One of the reasons why geothermal is so efficient is that the operating temperature difference between the evaporator and the condenser is low compared with other systems. This is more evident in commercial refrigeration systems, in which the temperature difference can go from a 0°F box to a 65°F earth rather than from a 0°F box to a 100°F rooftop. This reduced temperature differential is what makes geothermal more efficient than any other system. Even with future technological advances in equipment, the GHP system will still be more efficient, because these same enhanced components can be integrated into the geothermal system, which will improve their performance even more.

ORGANIZATIONS

States usually regulate drilling into the earth, and those regulations need to be known before beginning to design a

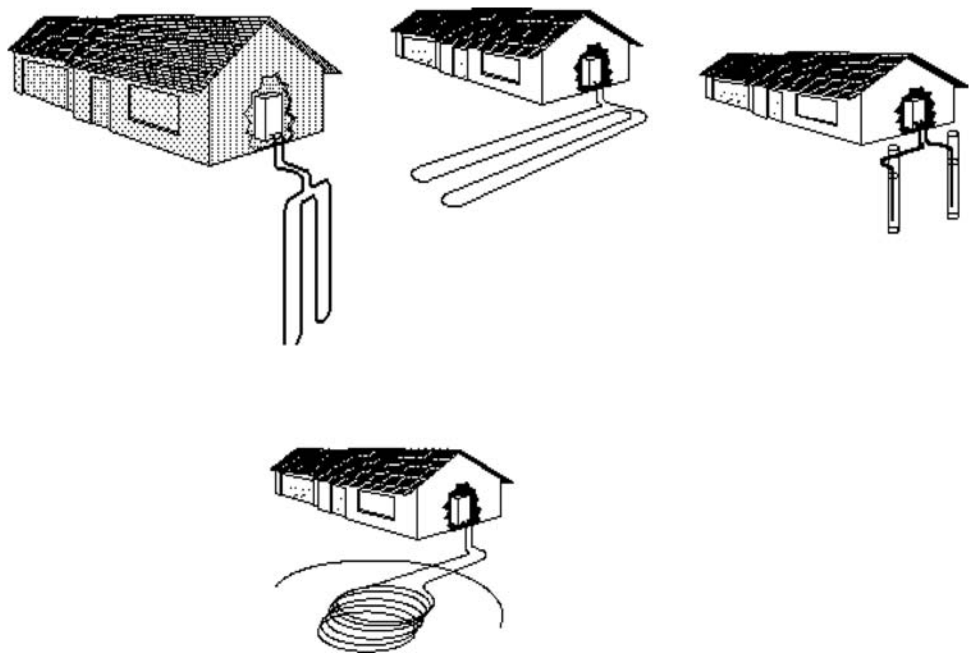


Fig. 3 Types of Earth heat exchangers.

geothermal system. Investigation should also be performed for local regulations that will impact the project. Sources of information about these systems can be found through the local utility companies, the International Ground Source Heat Pump Association (IGSHPA), the Geothermal Heat Pump Consortium (GHPC), and ASHRAE, to name just a few.

These systems have existed in Europe for many years, with significant research being done by Lund University in Sweden. In the United States, research has been performed by ASHRAE, IGSHPA (located at Oklahoma State University), the University of Alabama, and the Geo-Heat Center at the Oregon Institute of Technology. There are also several certified geothermal designers (CGD) in the United States who have been certified by the Association of Energy Engineers (AEE) to design these systems.

DEVELOPING A SYSTEM

The Building

In developing a GHP system, the standard terms used in the HVAC/R market still apply. The first phase in the design is the same with any HVAC system: begin by doing a load calculation for the system. A load calculation is a means of calculating how much heat is entering or leaving the space that is to be conditioned. It can be a space that people will occupy, such as a home or an office building, a refrigeration box for storing frozen foods, or heating the domestic water in a sports complex for people to shower. Anything that can be done with typical HVAC/R systems can be done with a geothermal system. There are all types of (GHP) systems for conditioning air or water for any type of service that is needed. Where geothermal systems differ in sizing from the typical systems is in the thermal mass, as discussed above.

At the beginning of the load calculation, a useage schedule needs to be assigned to every load that makes up the total load. Remember that a ton (kW) is a rate of flow and not a volume of heat, like ton-hour or kilowatt per hour. Here is an analogy for understanding this concept:

Think of designing a swimming pool. The first decision is how big to make the pool. An initial question would be how many people are going to be swimming at one time. The next things to determine are what the people will be doing in the pool and how much water will be forced out of the pool during the activity, because the water removed will have to be replaced. For the purpose of this analogy, the swimming pool will be filled using a garden hose.

As with typical load calculations, a coincident peak load (usually around July 21 or December 21) that represents the maximum heat transfer rate is needed to size the equipment. This is a rate of flow and does not represent the volume of heat being moved in either direction over time. Thus, a question emerges: how long will it take the garden hose to fill the pool?

The questions above are the same types of information needed to size a geothermal system. Apply this to another example calculation. The total equipment needed on the coincidental peak load days is 100 ton (351.6 kW). This load may exist for only a few hours of the 8760 h in each year, but the system may use 40 ton (140.6 kW) of the 100 ton for 1500 h of the year. The remaining hours of the 8760 h in the year will be at other partial-load conditions or even completely off. This is the information needed to figure out how much total heat is being put into a thermal mass that is going to absorb and store it for many days.

This principle of storing and reuse is what makes this technology a renewable energy. The heat that the system pulls out of the building in the summer will be held in the thermal mass until the next winter, when the system wants to pull the heat back out of the thermal mass and reuse it to heat the building. The object of the design is to make sure that the thermal mass is of sufficient size to store the heat annually or if it is imbalanced over the life of the system. The perfect system design is one in which the heat being put into the thermal mass in the summer is equal to the heat that needs to be pulled out in the winter. This is very rarely the case, however. In completing the calculation, one thing that is often missed is the internal heat of the machines performing the heat transfer. A 5-ton (17.5 kW), 10 EER machine, for example, can put 7 ton (24.6 kW) into the thermal mass but needs to extract only 3 ton (10.5 kW) from the thermal mass to produce its rated capacity. The 4 ton difference (7 − 3 ton) needs to be added into the equation. Therefore, the perfectly balanced system is not a balanced load calculation—there needs to be a higher heating requirement in the building than what needs to be rejected to the thermal mass.

The most important factor in a successful geothermal project is a thorough understanding of the heat to be transferred to and from the thermal mass. Typical software for doing these types of load calculations is readily available; programs include DOE2, TRACE 700, Energy Plus, and HAP.

The Thermal Mass

Once the heat volume is determined the next thing to examine is the thermal mass that will be used to transfer the heat to/from. If the system is to use the earth as the thermal mass, consultation with a professional geologist is recommended. A professional geologist can assist in understanding the thermal conductivity and deep-earth temperature for any site. There are also companies that will travel to the site, drill a borehole, and test the geological formation for these characteristics. The deep-earth average temperature is important because it is the starting point for transferring heat to and from the earth at this site. If a deep-earth average temperature is 50°F, for example, the cooling performance is going to be very good. On the other hand, if a deep-earth average temperature is 70°F, heating will produce better results. The thermal conductivity of the soils on each particular site will determine how fast heat moves through those particular soils in that particular borehole. Remember that the soil can vary widely from one site to the next and from one borehole to another, even at the same job site. There is no rule of thumb here, and it would be ill advised to try to apply one.

When the load profile is established and the thermal mass that is going to be used for heat storage is identified, the design can begin. Several programs available for sizing geothermal heat exchangers are GCHPcalc, GLHEPRO, GLCS, GL Design, GeoDeveloper, and GeoEase. Keep in mind that all the factors basically form a circular reference; no one factor is the driving force for the optimal design. All the variables interact and are dependent on one another for how they will react within the total system design of both the building and the thermal mass.

A Single Borehole

Another variable to recognize is borehole resistance. The thermal properties of the fluid that is carrying the heat to the thermal mass from the building equipment, and vice versa, should be known. The thermal properties of the thermal mass also should be known. Borehole resistance occurs between the fluid and the thermal mass as the heat passes from one to the other. The turbulence of the fluid, the type of pipe material, and the type of grout used in the borehole are examples of what creates the borehole resistance. The fluid must be turbulent, with a recommended Reynolds number between 2500 and 5000. Making the fluid more turbulent does not increase the heat transfer, but instead results in increased pumping energy to overcome friction losses in the pipe. Using too much pumping energy can negate the savings created by the GHP system.

The heat must be transferred from the fluid into the high-density polyethylene (HDPE) pipe that the loop is made of, then through the grout, and into the earth. Understanding the borehole resistance, how the borehole will be constructed by the contactor, and the equipment used to build the heat exchanger are important factors. Being familiar with the equipment that is readily available in the area or that may be used on the job site will need to be factored into the design. Requiring thermal conductivity tests during the construction process and verifying that the methods used in the construction process are within an acceptable range of the assumptions used in the design are parts of a good quality-assurance/quality-control protocol. There are software packages available for back-calculating the borehole resistance of a known construction method.

Grout is another very important factor in how the heat exchanger and borehole will perform. This is also one of the materials usually covered by state regulations on how a borehole is to be constructed. The grout is used to fill the annular space between the loop and the wall of the borehole created by the drill bit. Grout is one of the leading factors in borehole resistance, as discussed above. Although there is much debate about how to grout a borehole and what type of material should be used, it is an industry recommendation that this annular space be filled, and that the mixture used to fill it be determined during the design phase and be continuous throughout the construction process.

Some of the grouts typically used in geothermal heat exchangers are a 40% solids bentonite, a thermally enhanced bentonite, and a thermally enhanced cementitious-based grout. Some states may also allow the use of other materials, which cannot be covered within the limits of this entry. Investigation of all grouts available within the confines of the regulations for an area can be delivered by the contractors who may perform the work.

Fig. 4 is a cutaway diagram of a typical borehole. The loop is a long U-bend made out of HDPE.

OTHER CONSIDERATIONS

Another variable mentioned only slightly above is pumping. American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. has conducted several studies on the performance of pumping within geothermal systems. Investigation into pumping alternatives should be conducted so that enhanced energy savings can be achieved within the constraints of the design. Alternatives might include variable-speed pumping with the use of a variable-frequency drive on the pump motor or the use of distributive pumping techniques if the load

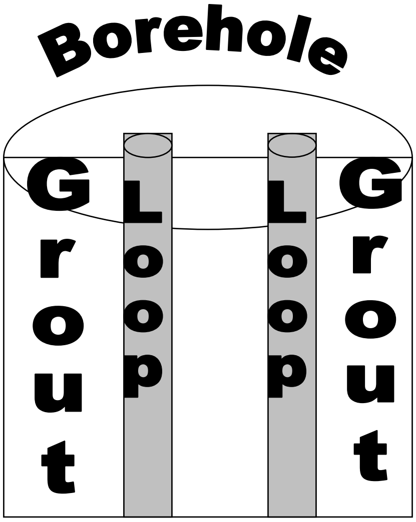


Fig. 4 Cutaway representation of a typical borehole.

profile indicates a high degree of diversity within the use of the facility.

One final factor in the design of the piping system is to remember that—as in any hydronic system—air is its worst enemy. These systems must be designed with purge and flush ports. These ports are a set of valves teed into the system to allow an external pump to be attached to the pipe. The minimum flow rate of the external pump must achieve a 2 ft/min flush rate to remove all air from the system. It is recommended that a fill line and an air-removal device be placed in the piping system. The properties of HDPE dictate that it have a large expansion coefficient, in that for every 10°F that the fluid temperature changes and for every 100 ft of pipe, the pipe will expand 1 ft. This expansion is a large volume that needs to be replaced by additional fluid if the system is very large. In residential systems that are very small, the designer may not want to consider a fill line, because it may create a water issue should a pump seal fail or another failure occur. Because HDPE is not 25/50-(flame spread/smoke developed) rated and should not be run in plenum spaces, the use of this pipe inside a building is restricted. The use of HDPE offers a considerable capital savings, however, if the building design can avoid running the pipe in plenum spaces.

CONCLUSION

Geothermal heat pump systems are the most efficient systems available today for many projects. This technology has been proved over time, and the tools necessary for a successful project exist in the market today. Design guides are available from ASHRAE (www.ashrae.org) and IGSHPA (www.igshpa.okstate.edu), as are other standards and specifications for these types of systems.

ACKNOWLEDGMENTS

I would like to express appreciation to Steve Redding of Redding Linden Burr Consulting Engineers, Fabian Nilsson of Electricity de France, and Catherine Tinkler, formerly of Lubbock Christian University, for reviewing this entry. Their comments were insightful and made this entry much easier to read.

REFERENCES

1. Tinkler, G.; Nilsson, F. Marketing brochure for EnLink Geoenergy Services, Inc., 2002.
2. Tinkler, G.; Nilsson, F. Lubbock Christian University case study, 2002.
3. U.S. Environmental Protection Agency. Available at: www.usctcgateway.net/tool. (accessed date on 8/4/2006).
4. American Society of Heating, Refrigerating, and Air-Conditioning Engineers. Available at: www.ashrae.org. (accessed date on 8/4/2006).
5. Geothermal Heat Pump Consortium. Available at: www.geoexchange.org. (accessed date on 8/4/2006).

BIBLIOGRAPHY

1. ASHRAE, IGSHPA, AEE, GHPC, Trane, and Carrier.

Global Climate Change

Amanda Staudt
Board on Atmospheric Sciences and Climate, The National Academies, Washington, D.C., U.S.A.

Nathan E. Hultman
Science, Technology, and International Affairs, Georgetown University, Washington, D.C., U.S.A.

Abstract

The global-mean temperature at Earth's surface has increased by about 0.6°C (1.1°F) over the past century. Most of this warming is due to the excess greenhouse gases emitted to the atmosphere by human activities such as fossil fuel use, agriculture practices, and land-use change. Many changes in natural systems have already resulted from this warming, including melting of glaciers and ice caps, increasing sea level, extended growing seasons, changes in precipitation regimes, and changes in the geographical distributions of plant and animal species. Current projections, based on reasonable assumptions of future energy-use practices, indicate that Earth will continue to warm during the 21st Century and beyond, in part because parts of the Earth system respond slowly to changes in greenhouse gas levels.

INTRODUCTION

For the past century, Earth's climate has been changing due to human activities. Observations show that Earth's surface warmed by approximately 0.6°C (1.1°F) on average in the 20th Century. Much of this warming has been attributed to increasing abundances of greenhouse gases emitted to the atmosphere by human activities, although it is difficult to quantify this contribution against the backdrop of natural variability and climate-forcing uncertainties. Atmospheric abundances of the major anthropogenic greenhouse gases (carbon dioxide; methane; nitrous oxide; halocarbons manufactured by humans, such as chlorofluorocarbons; and tropospheric ozone) reached their highest recorded levels at the end of the 20th Century, and all but methane have continued to rise. Major causes of this rise have been fossil fuel use, agriculture, and land-use change.

The emerging impacts of climate change on natural systems include melting glaciers and ice caps, the rising sea level, extended growing seasons, changes in precipitation regimes, and changes in the geographical distributions of plant and animal species. Additional impacts, to which it may be difficult for human and natural systems to adapt, could arise from events whose triggers are poorly understood. Human-induced global warming will continue during the 21st Century and beyond, because many parts of the Earth system respond slowly to changes in greenhouse gas levels and because altering established energy-use practices is difficult. Uncertainties remain about the magnitude and the impacts of future climate change, largely due to gaps in understanding of climate science and the socioeconomic drivers of climate change.

THE CLIMATE SYSTEM AND THE NATURAL GREENHOUSE EFFECT

While climate conventionally has been defined as the long-term statistics of the weather (e.g., temperature, cloudiness, precipitation), improved understanding of the atmosphere's interactions with the oceans, the cryosphere (ice-covered regions of the world), and the terrestrial and marine biospheres has led scientists to expand the definition of climate to encompass the oceanic and terrestrial spheres as well as chemical components of the atmosphere (Fig. 1). Physical processes within the atmosphere are influenced by ocean circulation, the reflectivity of Earth's surface, the chemical composition of the atmosphere, and vegetation patterns, among other factors.

The Sun provides almost all of Earth's energy. Solar radiation intercepted by Earth first encounters the atmosphere, which allows most of it to pass to Earth's surface. The intensity of radiation at the surface depends on the amount of incident radiation and on the orientation of the surface with respect to that radiation. The surface either reflects or absorbs this incoming radiation. Different surfaces reflect different amounts of sunlight. The fraction of solar energy reflected is defined as a surface's albedo. Albedos range from about 10% for open water, dark soil, and asphalt to about 80% for fresh snow.[1] Earth's average albedo is about 31%.

Keywords: Climate change; Global warming; Greenhouse; Carbon dioxide.

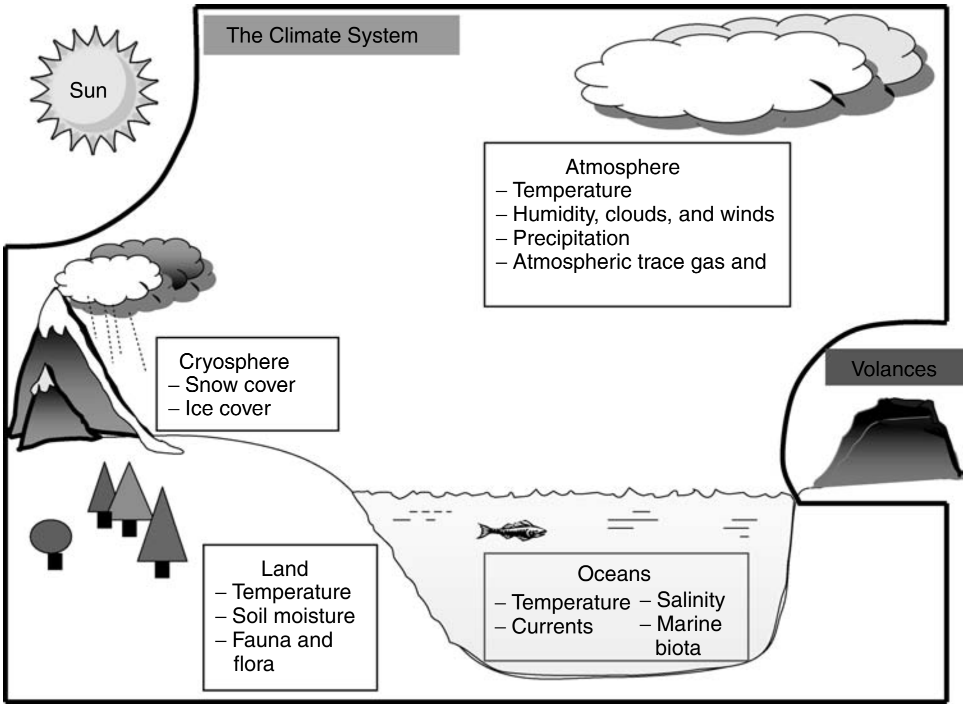


Fig. 1 The climate system.
Source: From National Academies Press (see Ref. 5).

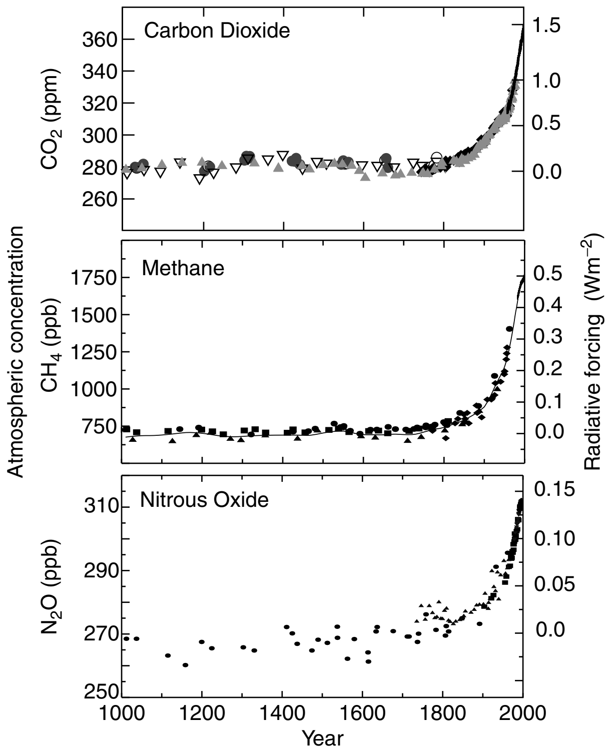


Fig. 2 Concentrations of major greenhouse gases retrieved from gas bubbles trapped in ice cores from Antarctica and Greenland.
Source: From Intergovernmental Panel on Climate Change (see Ref. 10).

Although we do not normally think of it as a radiative body, Earth—like all bodies with a nonzero temperature—emits electromagnetic radiation. For Earth's temperature, most of this radiation is in the form of infrared light. In the absence of an atmosphere, all the radiation emitted by Earth would escape to space. The balance of incoming solar radiation and outgoing infrared radiation would result in a global-mean temperature for Earth of 255 K (−18°C/0°F). However, some molecules in Earth's atmosphere absorb some of this outgoing infrared light, thereby increasing their temperature. These greenhouse gases in the atmosphere emit some energy back toward Earth, warming Earth's surface. This natural greenhouse effect, which is present in the absence of human activities, raises the global-mean surface temperature from 255 K to a comfortable 288 K (or about 15°C/59°F).[2]

Greenhouse gases that are present naturally in the atmosphere include water vapor (H_2O), carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and ozone (O_3). The most common greenhouse gas by quantity and the one exerting the greatest influence on the climate is water vapor; however, because water has a very short lifetime in the atmosphere (~1 week), any human perturbation will dissipate quickly. In most cases, the "greenhouse effect" or "climate change" refers not to this natural phenomenon but to additional, anthropogenic enhancements to the atmosphere's capacity to trap heat. Much higher concentrations of CO_2, CH_4, and N_2O have been observed in the past century than were naturally present for the past 1000 years (Fig. 2) and likely much

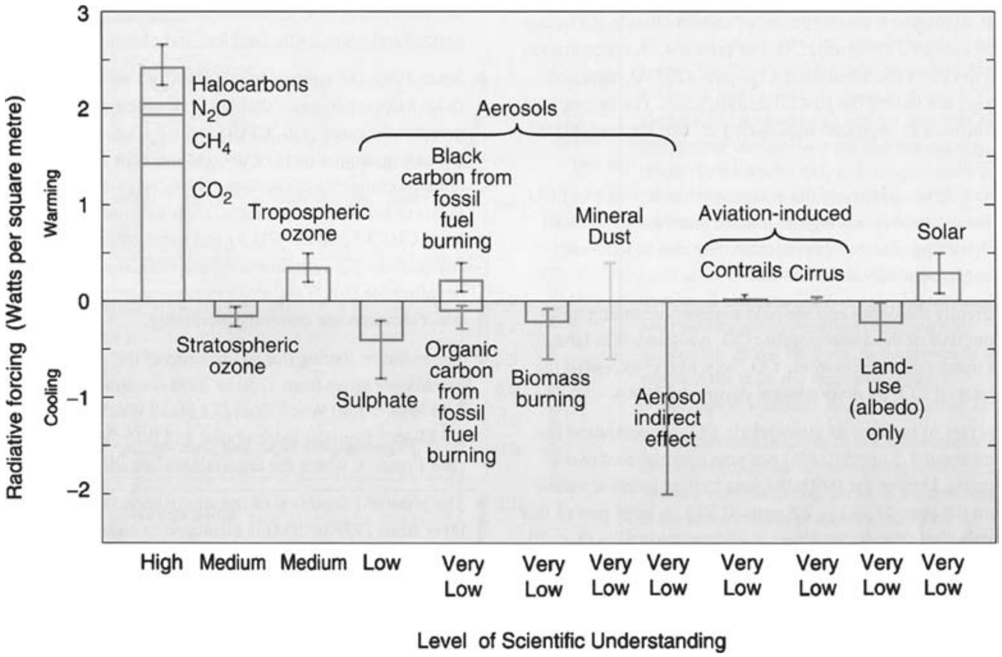


Fig. 3 Estimated radiative forcings since preindustrial times for Earth and the troposphere system. The height of the rectangular bar denotes a central or best estimate of the forcing, while each vertical line is an estimate of the uncertainty range associated with the forcing, guided by the spread in the published record and physical understanding, and with no statistical connotation. Each forcing agent is associated with a level of scientific understanding, which is based on an assessment of the nature of assumptions involved, the uncertainties prevailing about the processes that govern the forcing, and the resulting confidence in the numerical values of the estimate. On the vertical axis, the direction of expected surface temperature change due to each radiative forcing is indicated by the labels "warming" and "cooling."
Source: From Intergovernmental Panel on Climate Change (see Ref. 10).

longer.[3] Earth's surface is warmer now on average than it was at any time during the past 400 years, and it is likely warmer now than it was at any time in the past 2000 years.[4]

CLIMATE FORCINGS AND FEEDBACKS

Factors that affect climate change are usefully separated into forcings and feedbacks. Climate forcings are energy imbalances imposed on the climate system either externally or by human activities.[5] Examples include human-caused emissions of greenhouse gases, as discussed in the preceding section, as well as changes in solar energy input; volcanic emissions; deliberate land modification; or anthropogenic emissions of aerosols, which can absorb and scatter radiation. Climate forcings can be either direct or indirect. Direct radiative forcings are simple changes to the drivers of Earth's radiative balance. For example, added CO_2 absorbs and emits infrared radiation. Indirect radiative forcings create a radiative imbalance by first altering climate system components that lead to consequent changes in radiative fluxes; an example is the effect of aerosols on the precipitation efficiency of clouds. Fig. 3 provides a summary of the estimated contribution from major climate forcings. Additional information about specific climate forcings is provided in the discussion below.

Climate feedbacks are internal climate processes that amplify or dampen the climate response to an initial forcing.[6] An example is the increase in atmospheric water vapor that is triggered by an initial warming due to rising CO_2 concentrations, which then acts to amplify the warming through the greenhouse properties of water vapor (Fig. 4). Other climate feedbacks involve snow and ice cover, biogeochemistry, clouds, and ocean circulation. Some of the uncertainty about how the climate will change in the future stems from unresolved research questions on climate change feedbacks.

Natural Climate Forcings: Solar and Volcanic Variability

Variations in the Sun's activity and in Earth's orbital parameters cause natural forcing of climate. Radiometers on various spacecraft have been measuring the total solar irradiance since the late 1970s. There is an 11-year cycle in total solar irradiance of peak-to-peak amplitude ~1 W m^{-2} (0.1%) in the past three cycles. Allowing for reflection of 30% of this incident energy (Earth's albedo) and averaging over the globe, the corresponding climate forcing

Foss–Gre

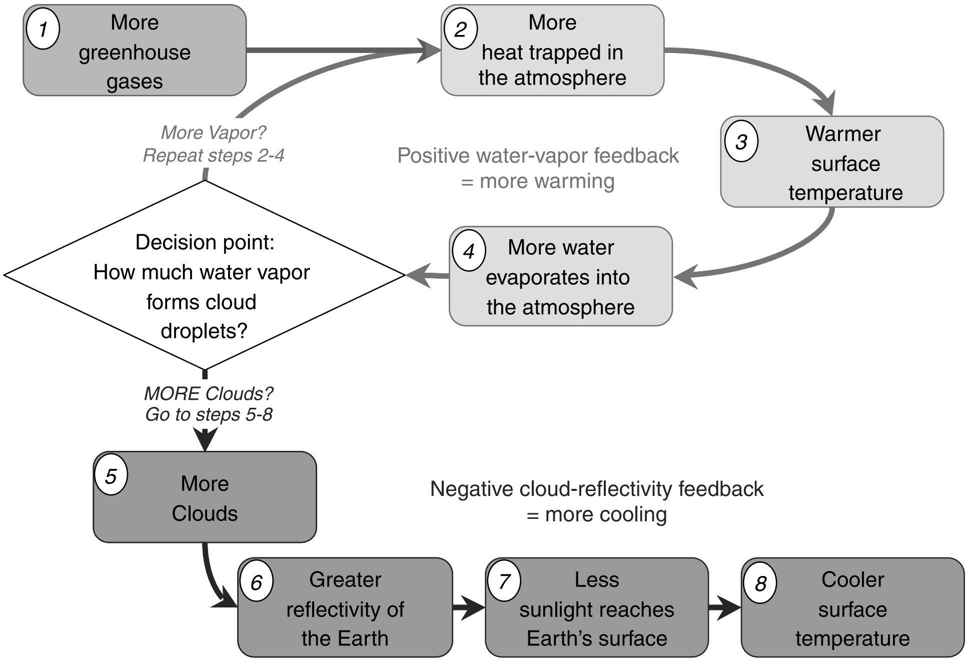


Fig. 4 This schematic illustrates just two out of the dozens of climate feedbacks identified by scientists. The warming created by emitting more greenhouse gases leads to evaporation of water into the atmosphere. But water itself is a greenhouse gas and can cause even more warming via positive water-vapor feedback. On the other hand, if the water vapor leads to more clouds, the result could be to counteract some of the warming because clouds can reflect incoming sunlight back to space. This chain of events is negative cloud-reflectivity feedback. Trying to understand whether water vapor will create more clouds and what kinds of clouds water vapor will create is a major research objective right now. The answer depends on weather patterns, where the evaporation takes place, and the amount of small soot particles suspended in the air.

is of order 0.2 W m^{-2}, although recent analyses have found little secular trend in solar irradiance over the past 30 years.[7] Knowledge of solar irradiance variations prior to 1979 is less certain, as it relies upon models of how sunspot and facular influences relate to solar irradiance observed since then. These models are used to extrapolate variations back to about 1610, when telescopes were first used to monitor sunspots. The amount of energy Earth receives from the Sun also depends on Earth's distance from the Sun, which does not remain constant. The eccentricity of Earth's orbit (currently 0.0167) and the tilt of its axis relative to the orbital plane result in continual changes to the amount and distribution of energy Earth receives. In modern times, this variation is $\pm 3.5\%$ during the year, with maximum energy and minimum distance in January.

Volcanic forcing has been the dominant source of natural global radiative forcing over the past millennium. Emissions from volcanic eruptions have multiple effects on climate, as listed in Table 1.[8] The greater prevalence of explosive volcanic activity during both the early and the late 20th Century and the dearth of eruptions from 1915 to 1960 represent a significant natural radiative forcing of 20th-Century climate.[9] Similarly, longer-term volcanic radiative forcing has been associated with a significant long-term forced cooling from 1000 to 1900, resulting from a general increase in explosive volcanic activity in later centuries.

Greenhouse Gas Forcing

The role of greenhouse gases in the climate system is well understood by scientists because instruments can accurately measure the abundances of these gases in the atmosphere and their radiative properties. The concentrations of CO_2, CH_4, N_2O, various halocarbons, and O_3 have increased substantially since preindustrial times, and they are the greatest contributors to total anthropogenic radiative forcing.[10] Many of these greenhouse gases are emitted primarily as a byproduct of fossil fuel combustion.

For a given gas, the total amount of heat-trapping ability depends on the efficiency of heat trapping for a given unit of gas (i.e., radiative forcing), the number of units present in the atmosphere, and the average length of time a given unit spends in the atmosphere. While these three components are enough to characterize a single gas, the large number of gases has prompted the development of an index called the global warming potential (GWP), which represents the relative impact of a particular greenhouse gas on the atmosphere's radiative balance.[10] See Table 2 for some GWP calculations. As the standard

Table 1 Effects of large explosive volcanoes on weather and climate

Effect and mechanism	Begins	Duration
Reduction of diurnal cycle	Immediately	1–4 days
Blockage of shortwave and emission of longwave radiation		
Reduced tropical precipitation	1–3 months	3–6 months
Blockage of shortwave radiation, reduced evaporation		
Summer cooling of Northern Hemisphere tropics and subtropics	1–3 months	1–2 years
Blockage of shortwave radiation		
Reduced Sahel precipitation	1–3 months	1–2 years
Blockage of shortwave radiation, reduced land temperature, reduced evaporation		
Stratospheric warming	1–3 months	1–2 years
Stratospheric absorption of shortwave and longwave radiation		
Winter warming of Northern Hemisphere continents	6–18 months	1 or 2 winters
Stratospheric absorption of shortwave and longwave radiation, dynamics		
Global cooling	Immediately	1–3 years
Blockage of shortwave radiation		
Global cooling from multiple eruptions	Immediately	Up to decades
Blockage of shortwave radiation		
Ozone depletion, enhanced UV radiation	1 day	1–2 years
Dilution, heterogeneous chemistry on aerosols		

Source: Reproduced/modified by permission of the American Geophysical Union (see Ref. 8).

Foss–Gre

reference gas, CO_2 has a GWP of 1, by definition. Over a time horizon of 100 years, CH_4 and N_2O have GWPs of 23 and 296, respectively. In other words, 1 additional kg of CH_4 in the atmosphere absorbs as much radiation as 23 additional kg of CO_2. However, these numbers change if the time horizon shifts.[10] By allowing greenhouse gases to be compared directly, GWPs enable policies that can reduce total climate impact by addressing the least-cost abatement options first.[11,12]

Atmospheric Aerosol Forcing

Aerosols are small particles or liquid droplets suspended in the atmosphere. Aerosols both scatter and absorb radiation, representing a direct radiative forcing. Scattering generally dominates (except for black carbon particles) so that the net effect is of cooling. The average global mean of aerosol-direct forcing from fossil fuel combustion and biomass burning is in the range of -0.2 to -2.0 W m^{-2}.[10] This large range results from uncertainties in aerosol sources, composition, and properties used in different models. Recent advances in modeling and measurements have provided important constraints on the direct effect of aerosols on radiation.[13–15] Aerosols have several indirect effects on climate, all arising from their interaction with clouds—particularly from their roles as cloud condensation nuclei (CCN) and ice nuclei (Table 3).

Land-Use Change Forcing

Land-use changes include irrigation, urbanization, deforestation, desertification, reforestation, the grazing of domestic animals, and dryland farming. Each of these alterations in landscape produces changes in radiative forcing, both directly and indirectly.[16,17] Direct effects include the change of albedo and emissivity resulting from the different types of land covers. For example, the development of agriculture in tropical regions typically results in an increase of albedo from a low value of forest canopies (0.05–0.15) to a higher value of agricultural fields, such as pasture (0.15–0.20). The Intergovernmental Panel on Climate Change (IPCC)[10] reports the globally averaged forcing due to albedo change alone as -0.25 ± 0.25 W m^{-2}. Significant uncertainties remain in estimating the effect of land-use change on albedo because of the complexity of land surfaces (e.g., the type of vegetation, phenology, density of coverage, soil color).

Indirect effects of land-cover change on the net radiation include a variety of processes related to (1) the ability of the land cover to use the radiation absorbed at the ground surface for evaporation, transpiration, and sensible heat fluxes (the impact on these heat fluxes caused by changes in land cover is sometimes referred to as thermodynamic forcing); (2) the exchange of greenhouse and other trace gases between the surface and the atmosphere; (3) the emission of aerosols (e.g., from dust); and (4) the distribution and melting of snow and

Table 2 Radiative forcing characteristics of some major greenhouse gases

	Contribution to direct radiative forcing		Concentration in 1998	Lifetime (yrs)	Global warming potential (GWP) for different time horizons		
	W m^{-2}	%			20 yrs	100 yrs	500 yrs
Carbon dioxide (CO_2)	1.46	60	365 ppm	5–200[a]	1	1	1
Methane (CH_4)	0.48	20	1745 ppb	12.0	62	23	7
Nitrous oxide (N_2O)	0.15	6	314 ppb	114	275	296	156
Halocarbons and related compounds	0.34	14	—	0.3–3200	40–15100	12–22200	4–16300

Total direct radiative forcing uncertainty is approximately 10%. The abbreviations parts per million (ppm) and parts per billion (ppb) refer to the ratio of greenhouse gas molecules to molecules of dry air.
[a] No single lifetime can be defined for CO_2 because it is affected by multiple removal processes with different uptake rates.
Source: From Intergovernmental Panel on Climate Change (see Ref. 10).

ice.[5] These effects are not yet well characterized or quantified.

EVIDENCE OF HUMAN-INDUCED CLIMATE CHANGE

Because we do not have a "control Earth" against which to compare the effects of our current changing atmosphere, incontrovertibly linking human activities and observed climate change is difficult. Scientists therefore rely on multiple, overlapping evidence of changes and then compare observed patterns of change with what our scientific understanding indicates should happen under anthropogenic climate change. This two-stage concept of discovering changes in climate and linking them to human activity is called detection and attribution. Evidence used to detect climate change is summarized in this section, and the use of climate models for attribution is discussed in the following section.

One piece of evidence of global warming is an increase in surface temperature since the 1900s, with particularly rapid increases since the late 1970s (Fig. 5). This dataset caused some controversy when researchers discovered that readings taken near the surface of Earth with thermometers appeared to be higher than readings of the lower atmosphere taken by satellites from above. Subsequent studies concluded that the warming trend in global-mean surface temperature observations during the past 30 years is undoubtedly real and is substantially greater than the average rate of warming during the 20th Century.[18] Satellite-and balloon-based observations of middle-troposphere temperatures, after several revisions of the data, now compare reasonably with one another and with observations from surface stations, although some uncertainties remain.[19,20]

The ocean, which represents the largest reservoir of heat in the climate system, has warmed by about 0.118°C (0.212°F), averaged over the layer extending from the surface down to 700 m, from 1955 to 2003 (Fig. 6).[21] Approximately 84% of the total heating of Earth's system (oceans, atmosphere, continents, and cryosphere) over the past 40 years has gone into warming the oceans. Recent studies have shown that the observed heat storage in the oceans is what would be expected by a human-enhanced greenhouse effect. Indeed, increased ocean heat content accounts for most of the planetary energy imbalance (i.e., when Earth absorbs more energy from the Sun than it emits back to space) simulated by climate models.[22]

Changes in several other climate indicators have been observed over the past decades, providing a growing body of evidence consistent with a human impact on the climate. For example, reductions in snow and ice cover are one important indicator.[10] Satellite observations indicate that snow cover has decreased by about 10% since the 1960s, while spring and summer sea-ice extent in the

Table 3 Overview of the different aerosol indirect effects associated with clouds

Effect	Cloud type	Description	Sign of top of the atmosphere radiative forcing
First indirect aerosol effect (cloud albedo or Twomey effect)	All clouds	For the same cloud water or ice content more but smaller cloud particles reflect more solar radiation	Negative
Second indirect aerosol effect (cloud lifetime or Albrecht effect)	All clouds	Smaller cloud particles decrease the precipitation efficiency, thereby prolonging cloud lifetime	Negative
Semidirect effect	All clouds	Absorption of solar radiation by soot leads to evaporation of cloud particles	Positive
Glaciation indirect effect	Mixed-phase clouds	An increase in ice nuclei increases the precipitation efficiency	Positive
Thermodynamic effect	Mixed-phase clouds	Smaller cloud droplets inhibit freezing, causing super cooled droplets to extend to colder temperatures	Unknown
Surface energy budget effect	All clouds	The aerosol-induced increase in cloud optical thickness decreases the amount of solar radiation reaching the surface, changing the surface energy budget	Negative

Source: From National Academies Press (see Ref. 5).

Northern Hemisphere has decreased by about 10%–15% since the 1950s. The shrinking of mountain glaciers in many nonpolar regions has also been observed during the 20th Century.

ATTRIBUTION OF OBSERVED CLIMATE CHANGE TO HUMAN INFLUENCE

An important question in global climate change is to what extent the observed changes are caused by the emissions of greenhouse gases and other human activities. Climate models are used to study how the climate operates today, how it may have functioned differently in the past, and how it may evolve in response to forcings. Built using our best scientific knowledge of atmospheric, oceanic, terrestrial, and cryospheric processes, climate models and their components are extensively tested against the full suite of observations of current and past climate to verify that they simulate a realistic version of the climate. Discrepancies between models and observations provide indications that we need to improve understanding of

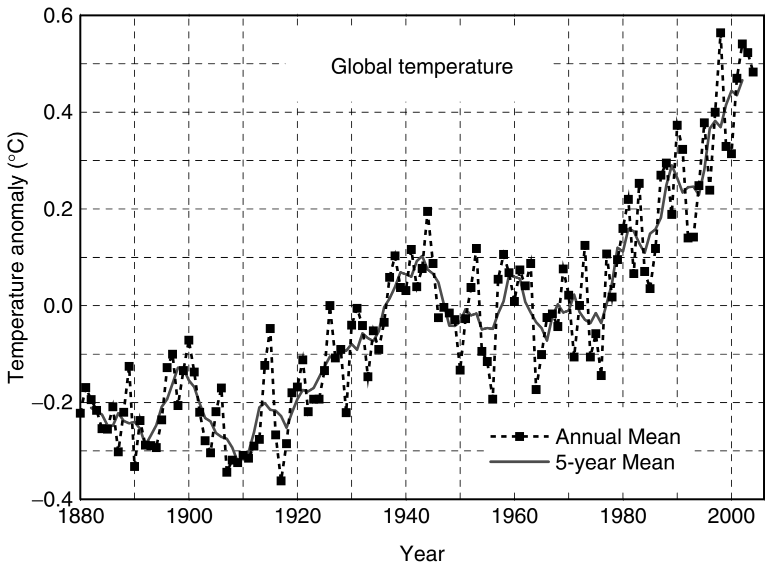


Fig. 5 Global annual-mean surface air temperature change derived from the meteorological station network. Data and plots are available from the Goddard Institute for Space Sciences (GISS) at http://data.giss.nasa.gov/gistemp/graphs.

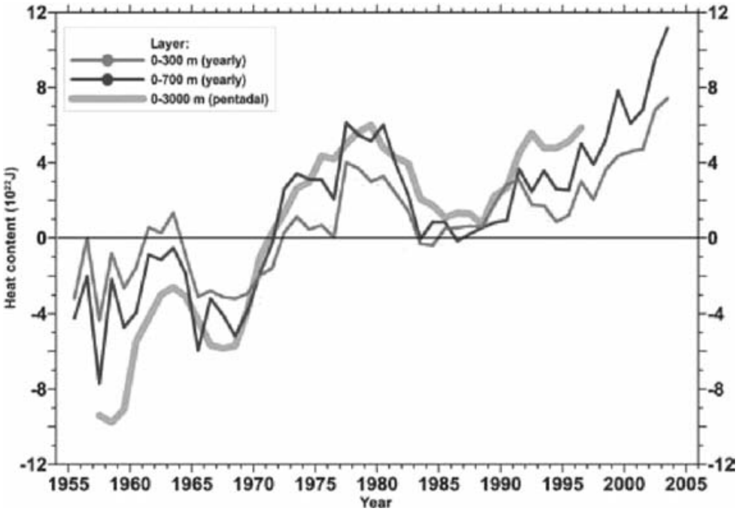


Fig. 6 Time series of (i) yearly ocean heat content (10^{22} J) for the 0–300 m and 0–700 m layers and (ii) 5-year running averages for 1955–1959 through 1994–1998 for the 0–3000 m layer.
Source: Reproduced/modified by permission of the American Geophysical Union (see Ref. 21).

physical processes, model representations of the processes, or in some cases the observations themselves. Hence, climate models contain our accumulated wisdom about the underlying scientific processes and can be no better than our observations of the system and our understanding of the climate.

Fig. 7 shows how scientists have used climate models to make the case that human activities have perturbed the climate since preindustrial times. In this experiment, the model is run with three different sets of climate forcings: (a) natural only, (b) anthropogenic only, and (c) natural and anthropogenic. When the natural or anthropogenic forcings are employed separately, the model is unable to reproduce the global-mean variation of temperature anomalies over the simulated time period. Only when both sets of forcings are used does the model capture the

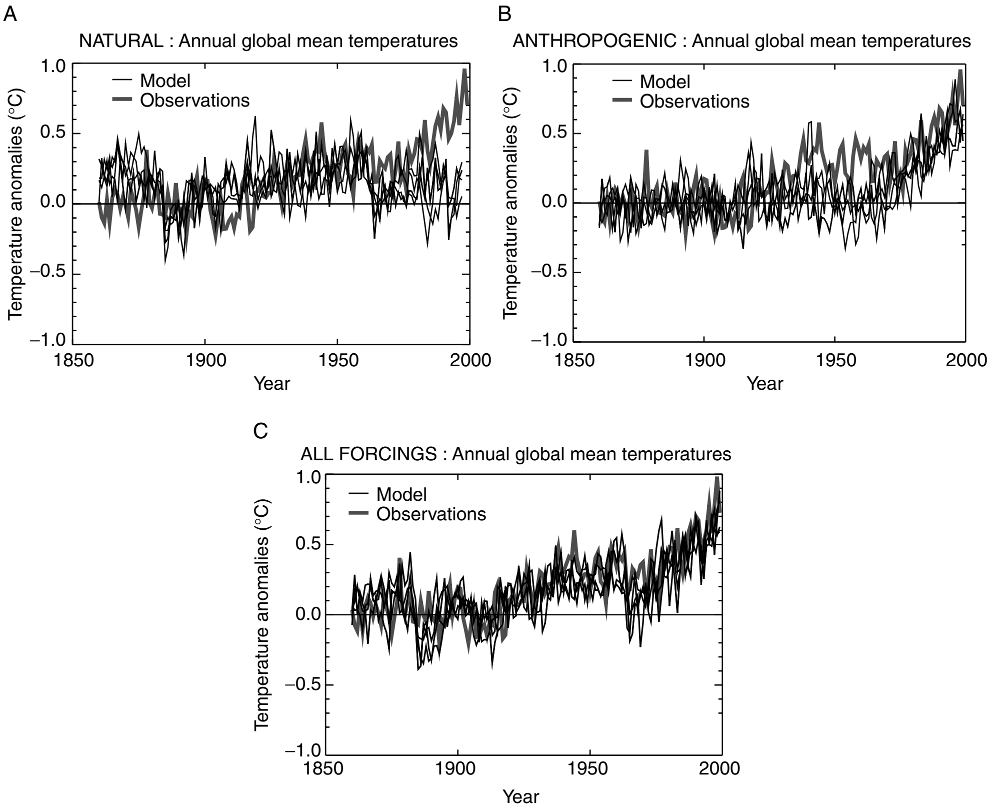


Fig. 7 Climate model results with (A) solar and volcanic forcings only; (B) anthropogenic forcings only; and (C) all forcings, both natural and anthropogenic.
Source: From Intergovernmental Panel on Climate Change (see Ref. 10).

nature of the variations, providing evidence that human activities have caused a significant fraction of warming in the past 150 years.[10]

PROJECTIONS FOR FUTURE CLIMATE CHANGE

The IPCC has concluded that by 2100, global surface temperatures will likely be from 1.4 to 5.8°C (2.5°F–10.4°F) above 1990 levels (Fig. 8) and that the combined effects of ice melting and seawater expansion from ocean warming will cause the global-mean sea level to rise by between 0.1 and 0.9 m.[10] Uncertainties remain about the magnitude and impacts of future climate change, largely due to gaps in understanding of climate science and the difficulty of predicting societal choices.

Climate changes in the coming century will not be uniformly distributed; some regions will experience more warming than others. There will be winners and losers from the impacts of climate change, even within a single region, but globally, the losses are expected to far outweigh the benefits. A changed climate will increase the likelihood of extreme heat and drought events.[23] High latitudes and polar regions are expected to see comparatively greater increases in average temperatures than lower latitudes, resulting in melting of permafrost and sea ice, which will result in additional costs for residents and in disruption to wildlife and ecosystems.[24] Precipitation changes, which are of great importance to agriculture, may have even more regional variability that is hard to predict.

Finally, several elements of Earth's system seem to be vulnerable to rapid destabilization. For example, the West Antarctic ice sheet and the Greenland ice sheet may be more prone to rapid melting than previously thought, and the loss of either of these would result in a large sea-level rise greater than 5 m. Moreover, the stability of the oceanic circulation that brings heat to Northern Europe has also been questioned. Because of feedback processes and the large uncertainty in system sensitivity, these outcomes are not easy to model and are usually not included in the gradual climate change projections quoted above.

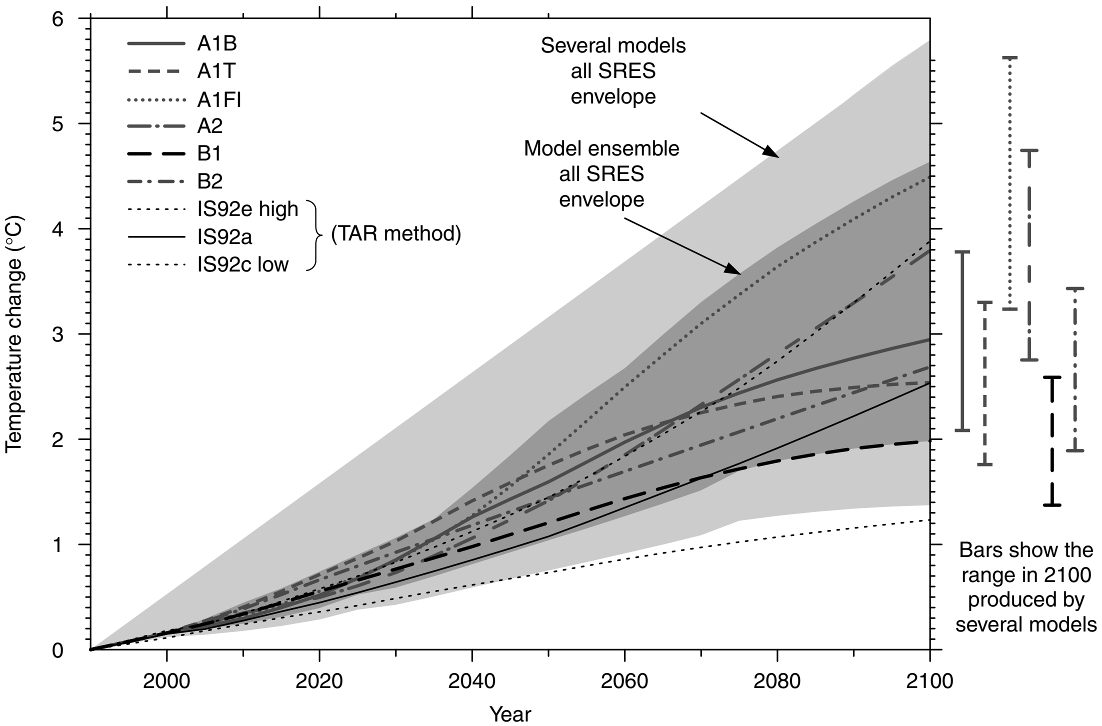


Fig. 8 Climate models are often used to simulate possible future climates to help inform decisions about policy responses to potential climate changes. This figure shows the range of plausible global-mean temperature change over the next 100 years, simulated by a collection of models. The spread in 2100 temperatures from 1.4 to 5.8°C reflects two factors: (1) Each model was run multiple times using different scenarios (indicated by different line textures) for future climate forcings, and (2) each model makes different assumptions about how the climate responds to those forcings. The scenarios range from those that assume continued acceleration of greenhouse gas emissions to those that assume more moderate growth or leveling off of emissions rates. SRES refers to the collection of scenarios presented in the Special Report on Emissions Scenarios. TAR refers to the IPCC Third Assessment Report.
Source: From Intergovernmental Panel on Climate Change (see Ref. 10).

Nevertheless, they are nontrivial threats and represent active areas of current research.

Unfortunately, the regions that will be most severely affected are often the regions that are the least able to adapt. Bangladesh, one of the poorest nations in the world, is projected to lose 17.5% of its land if sea level rises about 1 m (40 in.), displacing tens of thousands of people.[10] Several islands throughout the South Pacific and Indian Oceans will be at similar risk for increased flooding and vulnerability to storm surges. Although wetland and coastal areas of many developed nations—including the United States—are also threatened, wealthy countries may be more able to adapt to sea-level rise and threats to agriculture. Solutions could include limiting or changing construction codes in coastal zones and developing new agricultural technologies.

CONCLUSION

Research conducted to understand how the climate system may be changing—and in turn affecting other natural systems and human society—has led to significant advancement in scientific understanding, but many questions remain. Society faces increasing pressure to decide how best to respond to a changing climate and associated global and regional changes.

One way to address global climate change is to take steps to reduce the amount of greenhouse gases in the atmosphere. Because CO_2 and other greenhouse gases can remain in the atmosphere for many decades, the climate-change impacts from concentrations today will likely continue throughout the 21st Century and beyond. Failure to implement significant reductions in net greenhouse gas emissions now will make the job much harder in the future—both in terms of stabilizing CO_2 concentrations and in terms of experiencing more significant impacts. While no single solution can eliminate all future warming, many potentially cost-effective technological options could contribute to stabilizing greenhouse gas concentrations. These options range from personal choices such as driving less to national choices such as regulating emissions or seeking technologies to remove greenhouse gases from the atmosphere to international choices such as sharing energy technologies.

At the same time, it will be necessary to seek ways to adapt to the potential impacts of climate change. Climate is becoming increasingly important to public and private decision-making in various fields such as emergency management, water-quality assurance, insurance, irrigation and power production, and construction. For example, developing practical, "no regrets" strategies that could be used to reduce economic and ecological systems' vulnerabilities to change could provide benefits whether a significant climate change ultimately occurs or not. No-regrets measures could include low-cost steps to improve climate forecasting; to slow biodiversity loss; to improve water, land, and air quality; and to make institutions—such as the health care enterprise, financial markets, and transportation systems—more resilient to major disruptions.

REFERENCES

1. Houghton, J. *Global Warming: The Complete Briefing*, 3rd Ed.; Cambridge University Press: Cambridge, 2004; 351.
2. Houghton, J. *The Physics of Atmospheres*, 3rd Ed.; Cambridge University Press: Cambridge, U.K., 2002.
3. Petit, J.R.; Jouzel, J.; Raynaud, D.; et al. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature **1999**, *399*, 429–436.
4. NRC (National Research Council). *Surface Temperature Reconstructions for the Last 2000 Years*; National Academies Press: Washington, DC, 2006.
5. NRC (National Research Council). *Radiative Forcing of Climate Change: Expanding the Concept and Addressing Uncertainties*; National Academies Press: Washington, DC, 2005.
6. NRC (National Research Council). *Understanding Climate Change Feedbacks*; National Academies Press: Washington, DC, 2003.
7. Lean, J.; Rottman, G.; Harder, J.; et al. SORCE contributions to new understanding of global change and solar variability. Solar Phy. **2005**, *230*, 27–53.
8. Robock, A. Volcanic eruptions and climate. Rev. Geophys. **2000**, *38*, 191–219.
9. Crowley, T.J. Causes of climate change over the past 1000 years. Science **2000**, *289*, 270–277.
10. IPCC (Intergovernmental Panel on Climate Change). In *Climate Change 2001: The Scientific Basis*; Houghton, J.T., Ding, Y., Griggs, D.J., Eds.; Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press: Cambridge, U.K., 2001.
11. Prinn, R.; Jacoby, H.; Sokolov, A.; et al. Integrated global system model for climate policy assessment: Feedbacks and sensitivity studies. Climatic Change **1999**, *41* (3–4), 469–546.
12. Schneider, S.H.; Goulder, L.H. Achieving low-cost emissions targets. Nature **1997**, *389* (6646), 13–14.
13. Ramanathan, V.; Crutzen, P.J.; Kiehl, J.T.; et al. Aerosols, climate and the hydrological cycle. Science **2001**, *294*, 2119–2124.
14. Russell, L.M.; Seinfeld, J.H.; Flagan, R.C.; et al. Aerosol dynamics in ship tracks. J. Geophys. Res. Atmos. **1999**, *104* (D24), 31077–31095.
15. Conant, W.C.; Seinfeld, J.H.; Wang, J.; et al. A model for the radiative forcing during ACE-Asia derived from CIRPAS Twin Otter and R/V Ronald H. Brown data and comparison with observations. J. Geophys. Res. **2003**, *108* (D23), 8661.
16. Pitman, A.J. Review: The evolution of, and revolution in, land surface schemes designed for climate models. Int. J. Climatol. **2003**, *23*, 479–510.

17. Kabat, P., Claussen, M., Dirmeyer, P.A. et al., Eds.; *Vegetation, Water, Humans and the Climate—A New Perspective on an Interactive System*, Springer: Berlin, 2004.
18. NRC (National Research Council). *Reconciling Observations of Global Temperature Change*; National Academy Press: Washington, DC, 2000.
19. Mears, C.A.; Wentz, F.J. The effect of diurnal correction on satellite-derived lower tropospheric temperature. Science **2005**, *309*, 1548–1551.
20. Sherwood, S.C.; Lanzante, J.R.; Meyer, C.L. Radiosonde daytime biases and late-20th Century warming. Science **2005**, *309*, 1556–1559.
21. Levitus, S.; Antonov, J.; Boyer, T. Warming of the world ocean, 1955–2003. Geophys. Res. Lett. **2005**, *32* (L02604).
22. Hansen, J.; Nazarenko, L.; Ruedy, R.; et al. Earth's energy imbalance: Confirmation and implications. Science **2005**, *308*, 1431–1435.
23. Meehl, G.A.; Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st Century. Science **2004**, *305*, 994–997.
24. ACIA (Arctic Climate Impact Assessment). *Impacts of a Warming Arctic: Arctic Climate Impact Assessment*; Cambridge University Press: Cambridge, U.K., 2004.

Green Energy

Ibrahim Dincer
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology (UOIT), Oshawa, Ontario, Canada

Adnan Midilli
Department of Mechanical Engineering, Faculty of Engineering, Nigde University, Nigde, Turkey

Abstract
This book contribution discusses green energy for sustainable development and presents some key parameters and strategies to increase green-energy-based sustainability and the global peace level. The effects of technological, sectoral, and practical application impact ratios on green energy impact ratios and the green-energy-based sustainability ratio are also examined thoroughly. In addition, the key role of hydrogen, one of the most green energy carriers for the future, is discussed in terms of global unrest and global peace.

INTRODUCTION

Green energy can be considered a catalyst for energy security, sustainable development, and social, technological, industrial, and governmental innovations in a country. An increase in the green energy consumption of a country provides a positive impact on the economic as well as the social development of the country. Moreover, the supply and utilization of low-priced green fuel is particularly significant for global stability and peace because energy plays a vital role in industrial and technological developments around the world.

Critical energy issues in the 21st century will likely include energy security for almost 7 billion people, the expected global population by the middle of the 21st century, and global warming, mainly caused by CO_2 emissions generated from the combustion of fossil fuels.[1,2] Fossil fuels have caused some major problems for human health and human welfare due to their extensive use in various industrial and nonindustrial sectors. In reality, the main source of these problems is the extensive use of fossil-based technologies and strategies used by humans to govern throughout the centuries. In recent decades, the world has struggled with shortages of fossil fuels, pollution, and increased global energy requirements due to fast population growth, fast technological developments, and higher living standards. These factors have led to world population transition, migration, hunger, environmental (especially air and water pollution) problems, deteriorating health and disease, terrorism, energy and natural resources concerns, and wars. Also, problems with energy supply and use are related not only to global unrest, but also to such environmental concerns as air pollution, acid precipitation, ozone depletion, forest destruction, and the emission of radioactive substances. These issues must be taken into consideration simultaneously if humanity is to achieve a bright energy future with minimal environmental impact. Other environmental considerations have been given increasing attention by energy industries and the public. The concept that consumers share responsibility for pollution and its cost has been increasingly accepted. In some jurisdictions, the prices of many energy resources have increased over the last one to two decades, in part to account for environmental costs. Global demand for energy services is expected to increase by as much as an order of magnitude by 2050, while primary-energy demands are expected to increase by 1.5 ± 0.3 times.[2] Accordingly, humans have reached energy shortage and pollution levels that are not tolerable anymore. The urgent need in this regard is to develop green-energy-based permanent solutions for a sustainable future without any negative environmental and societal impacts. As a consequence, investigations for green energy should be encouraged particularly for green-energy-based sustainability and future world stability.

In this book contribution, the authors present key information on green-energy-based sustainability and global stability in accordance with major considerations that are presented in the following subtitles. In addition, this presents some key parameters like the green energy impact ratio and the green-energy-based sustainability ratio. Moreover, anticipated patterns of future energy use and consequent environmental impacts (focusing on acid precipitation, stratospheric ozone depletion, and the greenhouse effect) are comprehensively discussed. Also, potential solutions to current environmental problems are identified along with green energy technologies. The relationships between green energy and sustainable development are described. Throughout the article, several

Keywords: Green energy; Energy; Exergy; Environment; Sustainability; Parameters.

issues relating to green energy, the environment, and sustainable development are examined from both current and future perspectives. The specific objectives of this book contribution can be enumerated as follows:

- To help understand the main concepts and issues about green energy use and sustainability aspects
- To develop relationships between green energy use and sustainability development
- To encourage the strategic use and conservation of green energy sources
- To provide methods for energy security, implementation, and development
- To increase motivation on the implementation of green energy strategies for better energy supply
- To provide solutions to reduce negative environmental impacts by considering the possible green energy strategies
- To discuss possible green energy strategies for sectoral use

GREEN ENERGY

The most important property of green energy sources is environmental compatibility. In line with this characteristic, green energy sources will likely become the most attractive energy sources in the near future and will be the most promising energy sources for the technological and environmental perspectives of the 21st century, particularly in the context of sustainable development for the future. Green energy is one of the main factors that must be considered in discussions of sustainable development and future world stability. Several definitions of sustainable development have been put forth, including the following common one: "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."[3,4] There are many factors that can contribute to achieving sustainable development. One of them is the requirement for a supply of energy resources that is fully sustainable. A secure supply of green energy resources is generally agreed to be a necessary but not sufficient requirement for development within a society.

Furthermore, sustainable development within a society demands a sustainable supply of green energy sources and an effective and efficient utilization of green energy technologies. Green energy can be defined as an energy source that has zero or minimum environmental impact, is more environmentally benign and more sustainable, and is produced from solar, hydro, biomass, wind, and geothermal, etc. energies. These types of green energy reduce the negative effects of fossil energy resources and the overall emissions from electricity generation and decrease greenhouse gases. They give an opportunity to take an active role in improving the environment and meet clean energy demand for both industrial and nonindustrial applications. Considering the benefits of green energy, it can be said that the sustainability of green energy supply and progress is assumed to be a key element in the interactions between nature and society. In addition, sustainable development requires a supply of energy resources that is sustainably available at reasonable cost and causes no or minimal negative societal impacts. Clearly, energy resources such as fossil fuels are finite and thus lack the characteristics needed for sustainability while others such as green energy sources are sustainable over the relatively long term.[5] Particularly, low-priced green energy is the most essential means for increasing the sustainable technological development and industrial productivity as well as people's living standards in a society. Therefore, achieving solutions to the energy shortages and environmental problems presented today requires long-term potential actions for sustainable development. In this regard, green energy sources and technologies appear to be the one of the most efficient and effective solutions. It can be said that one solution to the impending energy shortage and environmental problems is to make much more use of green energy sources and technologies. Another important solution is to develop permanent and effective sustainable green energy strategies for the increase of sustainable global stability.[1,6] For these purposes, engineering practicality, reliability, applicability, economy, scarcity of supply, and public acceptability should also be considered.

GREEN ENERGY AND ENVIRONMENTAL CONSEQUENCES

Some of the global problems affecting world stability are presented in Fig. 1. Fossil fuel utilization effects such as global climate change, world energy conflicts, and energy source shortages have increasingly threatened world stability. These negative effects may be observed locally, regionally, and globally.

This concern arises due to world population growth, fast technological development, and increasing energy demand. In the past, fossil energy sources could be used to solve world energy problems. However, in this century, fossil fuels cannot continue indefinitely as the principal energy sources due to the rapid increase of world energy demand and energy consumption. Due to world population growth and the advance of technologies that depend on fossil fuels, the reserves of fossil fuels eventually will not be able to meet world energy demand. Energy experts point out the fact that reserves account for less than 40 years for petroleum, 60 years for natural gas, and 250 years for coal.[1] The increase of the energy consumption and energy demand indicates our dependence on the fossil fuels. If the increase of fossil fuel utilization continues in this manner, it is likely that the world will be affected by many negative problems due to fossil fuels. As one

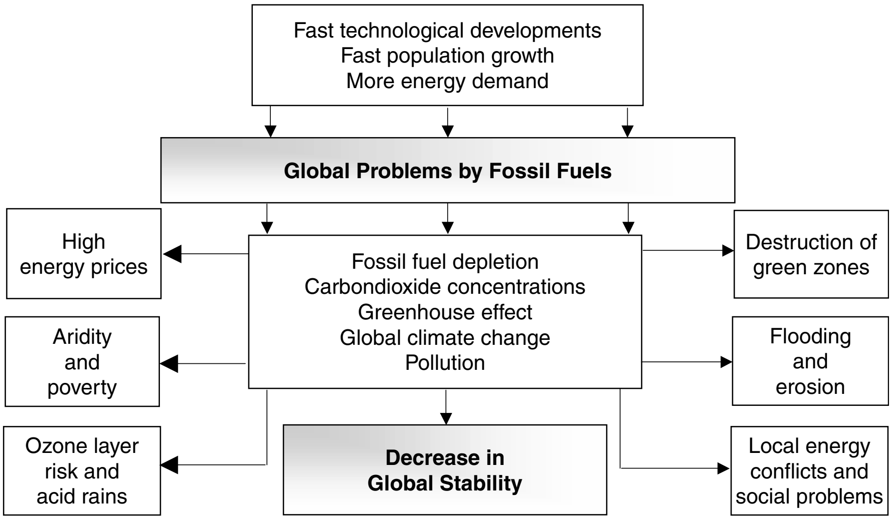


Fig. 1 Problems affecting global stability.

Foss—Gre

outcome, global stability will probably decrease. This effect is presented as flow chart in Fig. 2. For instance, shortages of fossil fuel reserves and global environmental problems will likely lead to global unrest throughout the world. As a result, local, regional, and world conflicts may appear across the world.

To further support these arguments, the observed and predicted consumptions of world primary energy, fossil fuels, and green energy from 1965 to 2050 are displayed in Fig. 3 based on the data taken from literature.[7,8]

According to Fig. 3, the quantities of world primary energy consumption are expected to reach 16502 Mtoe (Million tons of oil equivalent) by the year 2050. World population is now over 6 billion, double the population of 40 years ago, and it is likely to double again by the middle of the 21st century. The world's population is expected to rise to about 7.0 billion by 2010. Even if birth rates fall so that the world population becomes stable by 2050, the population will still be about 10 billion. Because the population is expected to increase drastically, conventional energy resource shortages are likely due to insufficient fossil fuel resource supplies. In the near future, therefore, green energy will become increasingly important to compensate for shortages of conventional resources. The correlations that have been applied here are as follows:

For world primary energy consumption :

$$\text{WPEC} = 143.57 \times Y - 277808 \quad (R^2 = 0.9902) \tag{1}$$

For world fossil fuel consumption :

$$\text{WFFC} = 113.61 \times Y - 219092 \quad (R^2 = 0.9968) \tag{2}$$

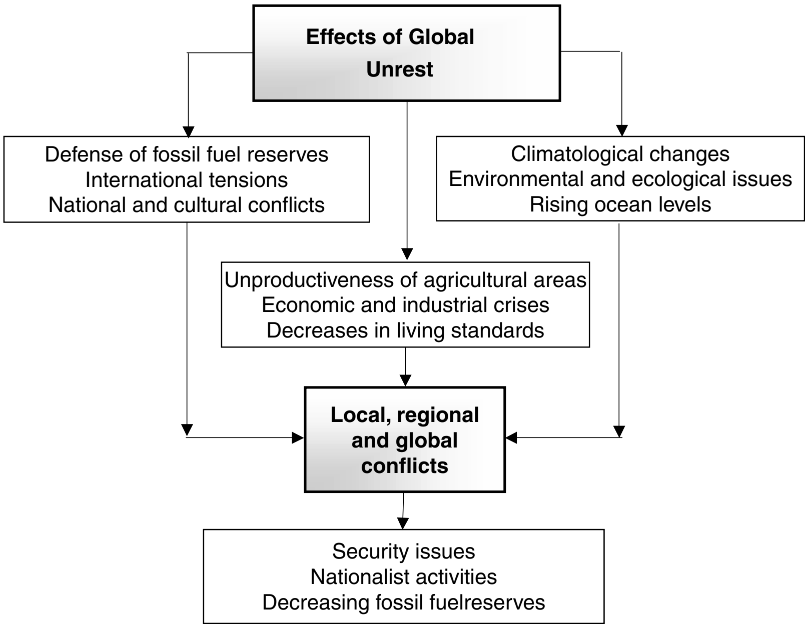


Fig. 2 Some possible effects and results of the global unrest.

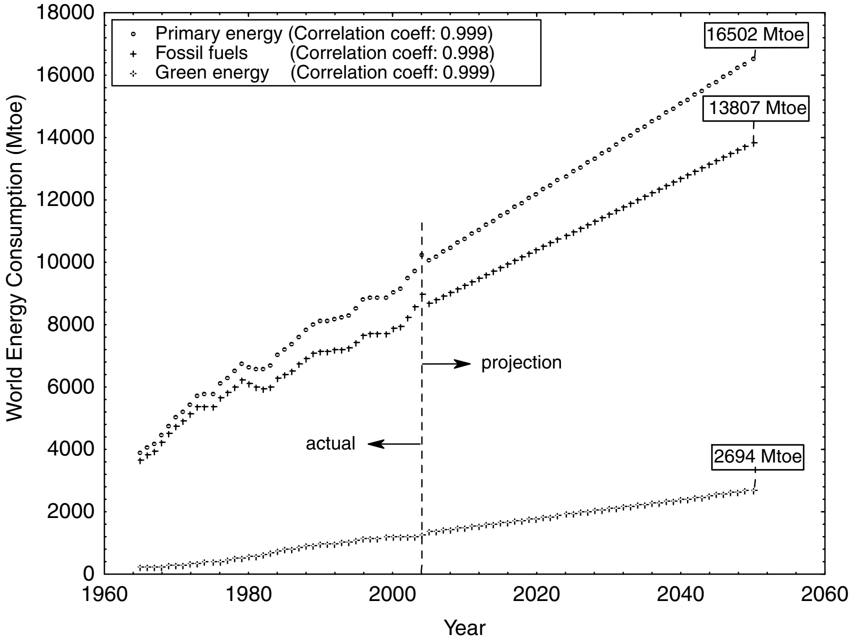


Fig. 3 Observed and projected quantities of world primary energy, fossil fuels, and green energy consumptions (Mtoe: million tons of oil equivalent).

For world green energy consumption :

$$\text{WGEC} = 29.956 \times Y - 58715 \quad (R^2 = 0.9985) \tag{3}$$

Here, WPEC is world primary energy consumption (Mtoe); WFFC, world fossil fuel consumption (Mtoe); WGEC, world green energy consumption; Y is time (years); and R is the correlation coefficient.

Energy shortages will accelerate the fluctuations of energy prices and economic recessions and decrease living standards and increase the unrest among countries. Decreased available fossil fuel reserves and increased fuel costs since the middle to late 1900s have led to variations in lifestyles and standards of life. These effects have, in some regions, decreased living standards of entire societies. Countries that need more energy resources have been purchasing cheaper energy sources. Countries that look after the future welfare of their societies have received the attention of countries that possess fossil fuel supplies, posing the potential for world conflict. Problems are often attributed to decreases in fossil fuel energy reserves. Those who seek a clean world must find appropriate alternatives to fossil fuels. Why not green energy? Green energy that is abundantly available all over the world can help:

- Reduce or stop conflicts among countries regarding energy reserves
- Facilitate or necessitate the development of new technologies
- Reduce air, water, and land pollution and the loss of forests
- Reduce illnesses and deaths caused by direct or indirect use of energy

Therefore, green energy and related technologies are needed to improve global stability by reducing the harmful effects of fossil-based energy consumption. Thus, the importance of green energy in reducing world problems and achieving a sustainable energy system should be emphasized and a transition to the green economy should be encouraged.

ENERGY, ENVIRONMENT, AND SUSTAINABILITY

Environmental concerns are an important factor in sustainable development. For a variety of reasons, activities that continually degrade the environment are not sustainable over time, e.g., the cumulative impact on the environment of such activities often leads over time to a variety of health, ecological, and other problems. A large portion of the environmental impact in a society is associated with its utilization of energy resources. Ideally, a society seeking sustainable development utilizes only energy resources that release no or minimal emissions to the environment and thus cause no or little environmental impact. However, because all energy resources lead to some environmental impact, it is reasonable to suggest that some (not all) of the concerns regarding the limitations imposed on sustainable development by environmental emissions and their negative impacts can be in part overcome through increased energy efficiency. Clearly, a strong relation exists between energy efficiency and environmental impact because for the same services or products less resource utilization and pollution is normally associated with increased energy efficiency. While not all green energy resources are inherently clean, there is such a diversity of choices that a shift to renewables carried out in the context of sustainable development could provide a far cleaner system than would be feasible by tightening controls on conventional energy.[9] Furthermore, being by nature site-specific, they favor power system decentralization and locally applicable solutions more or less independent of the national network. It enables citizens to perceive positive and negative externalities of energy consumption. Consequently, the small scale of the equipment often makes the time required from initial design to operation short, providing greater adaptability in responding to unpredictable growth and changes in energy demand. In this regard, sustainability often leads local and national authorities to incorporate environmental considerations into energy planning. The need to satisfy basic human needs and aspirations combined with the increasing world population will make the need for successful implementation of sustainable development increasingly apparent. Various criteria that are essential to achieving sustainable development in a society are as follows[3,4,10]:

- Information about and public awareness of the benefits of sustainability investments, environmental education, and training
- Appropriate energy strategies
- The availability of green energy resources and technologies
- A reasonable supply of financing
- Monitoring and evaluation tools

Environmental concerns are significantly linked to sustainable development. Activities which continually degrade the environment are not sustainable. For example, the cumulative impact on the environment of such activities often leads over time to a variety of health, ecological, and other problems. Clearly, a strong relation exists between efficiency and environmental impact because for the same services or products, less resource utilization and pollution is normally associated with increased efficiency.[3] Improved energy efficiency leads to reduced energy losses. Most efficiency improvements produce direct environmental benefits in two ways. First, operating energy input requirements are reduced per unit of output and pollutants generated are correspondingly reduced. Second, consideration of the entire life cycle for energy resources and technologies suggests that improved efficiency reduces environmental impact during most stages of the life cycle. In recent years, the increased acknowledgment of humankind's interdependence with the environment has been embraced in the concept of sustainable development. With energy constituting a basic necessity for maintaining and improving standards of living throughout the world, the widespread use of fossil fuels may have impacted the planet in ways far more significant than first thought. In addition to the manageable impacts of mining and drilling for fossil fuels and discharging wastes from processing and refining operations, the greenhouse gases created by burning these fuels is regarded as a major contributor to a global warming threat. Global warming and large-scale climate change have implications for food chain disruption, flooding, and severe weather events.

Therefore, use of green energy sources can help reduce environmental damage and achieve sustainability. The attributes of green energy technologies (e.g., modularity, flexibility, low operating costs) differ considerably from those for traditional, fossil-fuel-based energy technologies (e.g., large capital investments, long implementation lead times, operating cost uncertainties regarding future fuel costs). Green energy technologies can provide cost-effective and environmentally beneficial alternatives to conventional energy systems. Some of the benefits of the green energy-based systems are as follow[3,10]:

- They are relatively independent of the cost of oil and other fossil fuels, which are projected to rise significantly over time. Thus, cost estimates can be made reliably for green energy systems and they can help reduce the depletion of the world's nongreen energy resources.
- Implementation is relatively straightforward.
- They normally cause minimum environmental degradation and so they can help resolve major environmental problems. Widespread use of green energy systems would certainly reduce pollution levels.
- They are often advantageous in developing countries. In fact, the market demand for renewable energy technologies in developing nations will likely grow as they seek a better standard of living.

Under these considerations, green energy resources have some characteristics that lead to problematic but often solvable technical and economic challenges:

- Generally diffuse
- Not fully accessible
- Sometimes intermittent
- Regionally variable

The overall benefits of green energy technologies are often not well understood, leading to such technologies often being assessed as less cost-effective than traditional technologies. For green energy technologies to be assessed comprehensively, all of their benefits must be considered. For example, many green energy technologies can provide, with short lead times, small incremental capacity additions to existing energy systems. Such power generation units usually provide more flexibility in incremental supply than large devices like nuclear power stations.

Foss–Gre

GREEN ENERGY AND SUSTAINABLE DEVELOPMENT

Sustainability has been called a key to the solution of current ecological, economic, and developmental problems by Dincer and Rosen[10] A secure supply of energy resources is generally agreed to be a necessary but not sufficient requirement for development within a society. Furthermore, sustainable development demands a sustainable supply of energy resources that, in the long term, is readily and sustainably available at reasonable cost and can be utilized for all required tasks without causing negative societal impacts. Supplies of such energy resources as fossil fuels (coal, oil, and natural gas) and uranium are generally acknowledged to be finite; other energy sources such as sunlight, wind and falling water are generally considered renewable and therefore sustainable over the relatively long term. Wastes (convertible to useful energy forms through, for example, waste-to-energy incineration facilities) and biomass fuels are also usually viewed as sustainable energy sources.

In general, the implications of these statements are numerous and depend on how sustainable is defined.[3] For sustainable development, green energy can play an important role for meeting energy requirements in both industrial and local applications. Therefore, development and utilization of green energy sources and technologies should be given a high priority for sustainable development in a country. The need for sustainable energy development is increasing rapidly in the world. Widespread use of green energy sources and technologies is important for achieving sustainability in the energy sectors in both developing and industrialized countries. Thus, green energy resources and technologies are a key component of sustainable development for three main reasons:

- They generally cause less environmental impact than other energy sources. The variety of green energy resources provides a flexible array of options for their use.
- They cannot be depleted. If used carefully in appropriate applications, green energy resources can provide a reliable and sustainable supply of energy almost indefinitely.
- They favor system decentralization and local solutions that are somewhat independent of the national network, thus enhancing the flexibility of the system and providing economic benefits to small isolated populations. Also, the small scale of the equipment often reduces the time required from initial design to operation, providing greater adaptability in responding to unpredictable growth and changes in energy demand.

Major considerations involved in the development of green energy technologies for sustainable development as modified from[10] are presented in Fig. 4.

As a consequence, it can be said that green energy and technologies are definitely needed for sustainable development that ensures a minimization of global unrest.

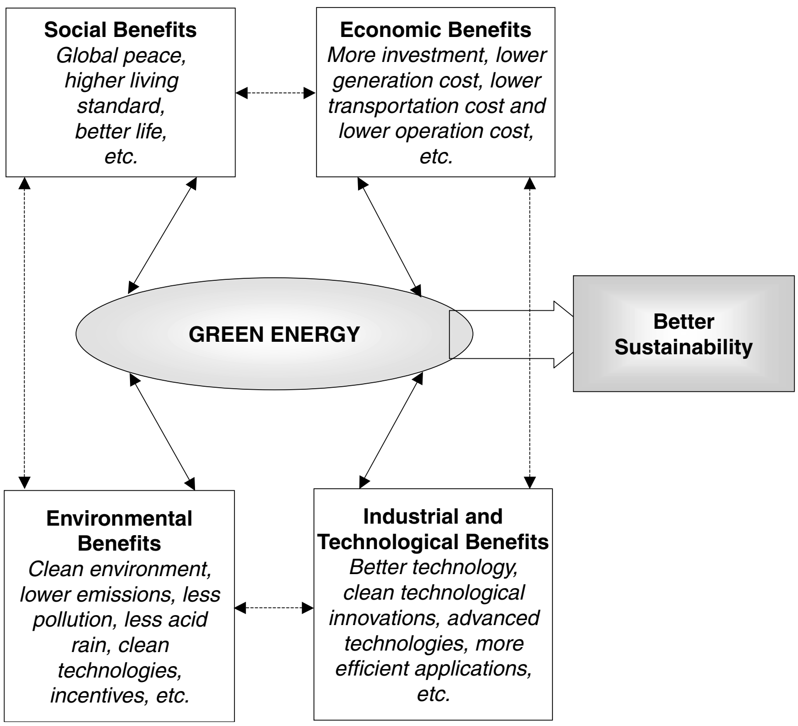


Fig. 4 Major considerations involved in the development of green energy technologies for sustainable development.

The relation between green energy and sustainability is of great significance to the developed countries as well as developing or less developed countries. Moreover, examining the relations between green energy sources and sustainability makes it clear that green technology is directly related to sustainable development. Therefore, attaining sustainable development requires that green energy resources be also used, and is assisted if resources are used efficiently.[4,10] Thus, if sustainable green energy technologies are effectively put into practice, the countries may maximize the benefits from the green energy sources and technologies while minimizing the global unrest[1] associated with the use of fossil fuel energy sources. It expected that this initiative can contribute to development over a longer period of time, i.e., to make development more sustainable.

GREEN ENERGY RESOURCES AND TECHNOLOGIES

Since the oil crises in the early 1970s, there has been active worldwide research and development in the field of green energy resources and technologies. During this time, energy conversion systems that were based on green energy technologies appeared to be most attractive because of facts such as the projected high cost of oil and the cost effectiveness estimates and easy implementation of green energy technologies. Furthermore, in more recent times, it has been realized that green energy sources and technologies can have a beneficial impact on essential technical, environmental, economic, and political issues of the world. As pointed out by Hartley,[3,9] green energy technologies produce marketable energy by converting natural phenomena into useful energy forms. These technologies use the energy inherent in sunlight and its direct and indirect impacts on the earth (photons, wind, falling water, heating effects, and plant growth), gravitational forces (the tides), and the heat of the earth's core (geothermal) as the resources from which they produce energy. These resources represent a massive energy potential which dwarfs that of equivalent fossil resources. Therefore, the magnitude of these is not a key constraint on energy production.

However, they are generally diffused and not fully accessible, some are intermittent, and all have distinct regional variabilities. Such aspects of their nature give rise to difficult, but solvable, technical, institutional, and economical challenges inherent in development and the use of green energy resources. Despite having such difficulties and challenges, the research and development on green energy resources and technologies has been expanded during the past two decades because of the facts listed above. Recently, significant progress has been made by (1) improving the collection and conversion efficiencies, (2) lowering the initial and maintenance costs, (3) increasing the reliability and applicability, and (4) understanding the phenomena of green energy technologies.

Green energy technologies become important as environmental concerns increase, utility (hydro) costs climb, and labor costs escalate.[3,9] The uncertain global economy is an additional factor. The situation may be turned around with an increase in research and development in the advanced technologies fields, some of which are closely associated with green energy technologies. This may lead to innovative products and job creation supported by the governments. The progress in other technologies, especially in advanced technologies, has induced some innovative ideas in green energy system designs. The ubiquitous computer has provided means for optimizing system performance, costs/benefits, and environmental impacts even before the engineering design was off the drawing board. The operating and financial attributes of green energy technologies, which include modularity and flexibility and low operating costs (suggesting relative cost certainty) are considerably different than those for traditional, fossil-based technologies, whose attributes include large capital investments, long implementation lead times, and operating cost uncertainties regarding future fuel costs. The overall benefits of green energy technologies are often not well understood and consequently they are often evaluated to be not as cost effective as traditional technologies. In order to comprehensively assess green energy technologies, however, some of their benefits that are often not considered must be accounted for. Green energy technologies, in general, are sometimes seen as direct substitutes for existing technologies so that their benefits and costs are conceived in terms of assessment methods developed for the existing technologies. Many government organizations and universities recognize the opportunity and support the efforts to exploit some commercial potential by:

- Analyzing opportunities for green energy and working in consultation with industry to identify research, development, and market strategies to meet technological goals.
- Conducting research and development in cooperation with industry to develop and commercialize technologies.
- Encouraging the application of green energy technologies to potential users, including utilities.
- Providing technical support and advice to industry associations and government programs that are encouraging the increased use of green energy. In order to realize the energy and the economic and environmental benefits that green energy sources offer, the following integrated set of activities should be acted on accordingly.[3,9,10]
- Conducting research and development. The priorities should be set in close consultation with industry to

reflect their needs. Most research is conducted through cost-shared agreements and falls within the short-to-medium term. Partners in research and development should include a variety of stakeholders in the energy industry such as private sector firms, utilities across the country, provincial governments, and other federal departments.

- Assessing technology. Data should be gathered in the lab and through field trials on factors such as cost benefit, reliability, environmental impact, safety, and opportunities for improvement. This data should also assist the preparation of technology status overviews and strategic plans for research and development.
- Developing standards. The development of technical and safety standards is needed to encourage the acceptance of proven technologies in the marketplace. Standards development should be conducted in co-operation with national and international standards writing organizations as well as other national and provincial regulatory bodies.
- Transferring technology. Research and development results should be transferred through the sponsorship of technical workshops, seminars, and conferences, as well as through the development of training manuals and design tools and the publication of technical reports. Such activities will also encourage potential users to consider the benefits of adopting green energy technologies. In support of developing near-term markets, a key technology transfer area is to accelerate the use of green energy technologies in a country's remote communities.

Such activities will also encourage potential users to consider the benefits of adopting renewable energy technologies. In support of developing near-term markets, a key technology transfer area is to accelerate the use of green energy technologies in a country's remote communities.

ESSENTIAL FACTORS FOR SUSTAINABLE GREEN ENERGY TECHNOLOGIES

There are various essential parameters, as outlined and detailed in Fig. 5. These factors can help in identifying and achieving required green energy strategies and technologies for sustainable development. As shown Fig. 5, green energy technologies are largely shaped by broad and powerful trends that are rooted in basic human needs. In conjunction with this, the increasing world population requires the definition and successful implementation of green energy technologies. Briefly, the important parameters and their interrelations, as outlined in Fig. 5, are definitely required to carry out the best green energy

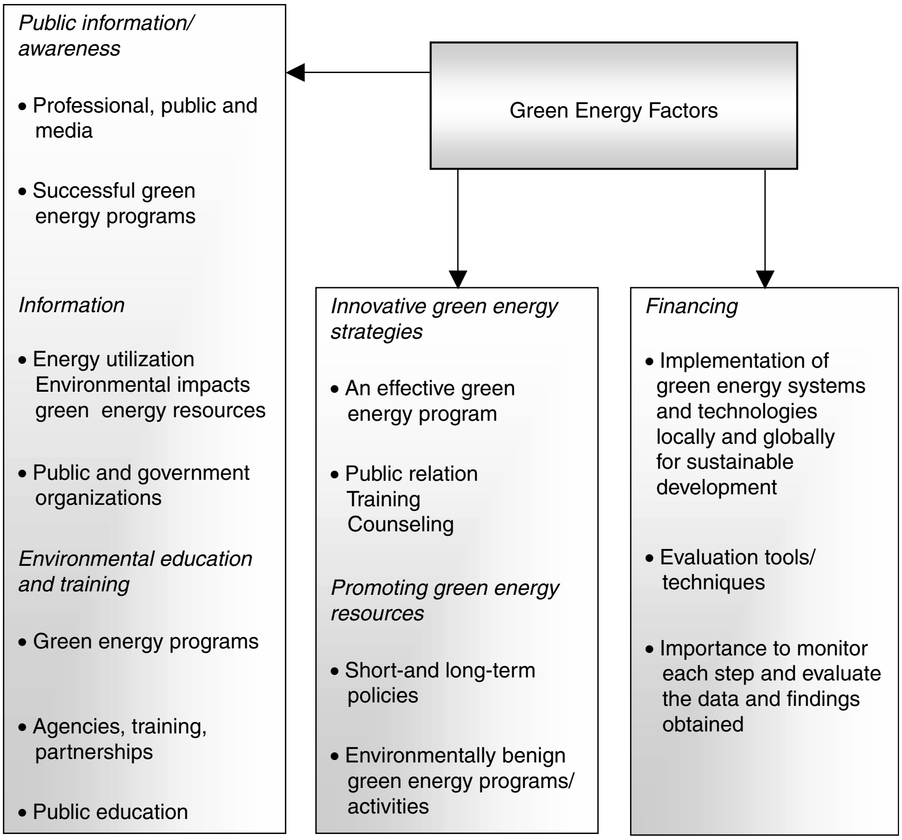


Fig. 5 Essential factors for green energy technologies.

program and to select the most appropriate green energy technologies for sustainable development.

EXERGETIC ASPECTS OF GREEN ENERGY TECHNOLOGIES

The impact of energy resource utilization on the environment and the achievement of increased resource-utilization efficiency are best addressed by considering exergy. The exergy of an energy form or a substance is a measure of its usefulness, quality, or potential to cause change and provide the basis for an effective measure of the potential of a substance or energy form to impact the environment. In practice, a thorough understanding of exergy and the insights it can provide into the efficiency, environmental impact, and sustainability of green energy technologies are helpful if not required for the engineer or scientist working in the area of green-energy-based environmental systems.[4,11] For green energy technologies, applications of exergy methods can have numerous broad and comprehensive benefits:

- A better knowledge of the efficiencies and losses for the technologies and systems and how they behave and perform
- A clearer appreciation of the environmental impacts of green energy technologies as well as the mitigation of environmental that they can facilitate
- Better identification of the ways green energy technologies can contribute to sustainable development

GREEN ENERGY APPLICATIONS

Green energy technologies are expected to play a key role in sustainable energy scenarios for the future. The foremost factor that will determine the specific role of green energy and technologies will likely be energy demand. Therefore, in order to compensate the energy requirement, it will be possible to produce green energy from renewable energy sources such as hydraulic, solar, wind, geothermal, wave and biomass, etc. If so, the green energy and technologies can be utilized for many application fields as shown in Fig. 6.

Thus, it can be said that green energy and technologies, which are abundantly available, can help:

- Provide more environmentally benign and more sustainable future
- Increase energy security
- Facilitate or necessitate the development of new, clean technologies
- Reduce air, water, and soil pollution and the loss of forests
- Reduce energy-related illnesses and deaths
- Reduce or stop conflicts among countries regarding energy reserves

Therefore, green energy and related technologies are needed to ensure global stability by reducing the harmful effects of fossil-based energy consumption. Thus, the importance of green energy in reducing the world problems and achieving a sustainable energy system should be emphasized considering the sustainable green energy strategies; and a transition to green energy economy should be encouraged and developed countries in particular should increase investments in green energy and technologies.

GREEN ENERGY ANALYSIS

In light of the above major considerations, in order to accelerate the use of green energy sources and technologies and the implementation of green energy strategies, and in this regard, to describe the relationship between the global stability and sustainable development, some key steps are presented as follows:

- Key strategies
- Green energy based sustainability ratio
- Global unrest and peace

Green-Energy-Based Sustainable Development Ratio

In order to discuss the key role of green energy for sustainable development and global stability, the general algebraic form of the equation that is the green-energy-based sustainable development ratio should be presented based on the above strategies. For this purpose, the works that were early presented by Midilli et al.[1,2,6,12] and Dincer[3,10] were taken as the reference basis. The following important parameters, which are sectoral, technological, and practical application impact ratios, are taken into consideration to estimate the green-energy-based sustainable development ratio. The detailed derivations of these parameters are presented in the literature.[2] Briefly, the green-energy-based sustainable development ratio is mainly based on the following parameters:

- *Sectoral impact ratio*, (R_{si}) (ranging from 1 to 1/3), is based on the provided financial support ($C_{p,si}$) of public, private, and media sectors for transition to green-energy-based technologies, and depending on the total green energy financial budget (C_{geb}), as a reference parameter.[2]
- *Technological impact ratio*, (R_{ti}) (ranging from 1 to 1/3), is based on the provided financial support ($C_{p,ti}$)

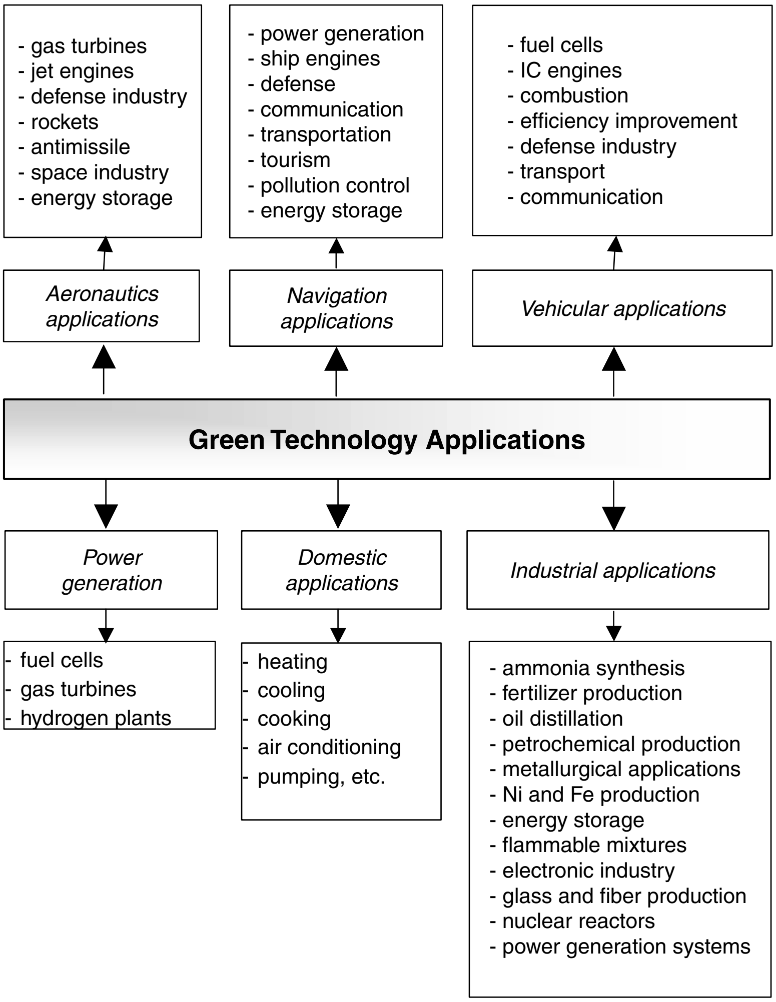


Fig. 6 Possible application fields of green energy and technologies for sustainable development.

for research and development, security, and analysis of green-energy-based technologies, and also depending on the total green energy financial budget (C_{geb}) as a reference parameter.[2]

- *Practical application impact ratio*, (R_{pai}) (ranging from 1 to 1/3), is based on the provided financial support ($C_{p,pai}$) for projection, production, conversion, marketing, distribution, management, and consumption of green fuel from green energy sources, and also depending on the total green energy financial budget (C_{geb}), as a reference parameter.[2] Here, it should be emphasized that these parameters can be defined as the ratio of the provided financial support to the total green energy financial budget in a country. In addition, it should be always remembered that it is assumed that the financial share of each parameter is equal to one-third of the total green energy financial budget in a country. Utilization ratio of green energy depends on the green energy utilization and the world primary energy quantity. Green-energy-based global stability is a function of utilization ratios of coal, petroleum, and natural gas, and utilization ratios of green energy at certain utilization ratios of these fuels, and utilization ratio of green energy. Therefore, the green-energy-based sustainable development ratio is written in algebraic form as follows[2]:

$$R_{ges} = R_{gei} \times R_{geu} \tag{4}$$

where
$R_{gei} = (R_{si} + R_{ti} + R_{pai})$, green energy impact ratio;
$R_{si} = C_{p,\ si}/C_{geb}$, sectoral impact ratio, which is estimated based on the provided financial support of the public, private, and media sectors for transition to green energy ($C_{p,si}$) and also the green energy budget allocation of a country (C_{geb});
$R_{ti} = C_{p,\ ti}/C_{geb}$, technological impact ratio, which is estimated based on the provided financial supports for research and development, security, and analysis of green-energy-based technologies ($C_{p,ti}$) and also the green energy budget allocation of a country (C_{geb});
$R_{pai} = C_{p,\ pai}/C_{geb}$, practical application impact ratio, which is estimated based on the provided financial supports for projection, production, conversion,

marketing, distribution, management, and consumption of green fuel from green energy sources, and also the green energy budget allocation of a country (C_{geb});

$R_{geu} = 1 - Q_{wffc}/Q_{wpec}$, green energy utilization ratio, which is also defined as a function of fossil fuel utilization ratio (R_{ffu}).

Here Q_{wffc} explains world fossil fuel consumption (M) and Q_{wpec} world primary energy consumption (M).

Global Unrest and Peace

Fossil fuels such as petroleum, coal, and natural gas, which have been extensively utilized in industrial and domestic applications for a long time, have often been the cause of global destabilization and unrest. This problem is likely to increase in significance in the future and suggests the need for investigations of, among other factors, the role of hydrogen energy relative to future global unrest and global peace. In general, the global unrest arising from the use of fossil fuels is considered as a function of the usage ratios of these fossil fuels.[12] In order to estimate the level of global unrest, it is important to select a proper reference value. Therefore, it is assumed that the lowest value of global unrest and the highest value of global peace are equal to 1, which is a reference point to evaluate the interactions between global unrest and global peace.[12] Consequently, the general algebraic case form of global unrest expression is as follows[12]:

$$GU = \left\{ 1 - \left[\left(\frac{q_p}{q_{wpec}} + \frac{q_c}{q_{wpec}} + \frac{q_{ng}}{q_{wpec}} \right) - (r_{H_2-p} + r_{H_2-c} + r_{H_2-ng}) \right] \right\}^{-1} \quad (5)$$

where q_p, q_c and q_{ng} define the consumption quantities of petroleum, coal, and natural gas, respectively; q_{wpec} defines the quantity of world primary energy consumption; r_{H_2-p} is the utilization ratio of hydrogen from nonfossil fuels at a certain utilization ratio of petroleum; r_{H_2-c} is the utilization of hydrogen from nonfossil fuels at a certain utilization ratio of coal; r_{H_2-ng} is the utilization of hydrogen from nonfossil fuels at a certain utilization ratio of natural gas. The values of q_p, q_c, q_{ng}, and q_{wpec} can be taken from the literature.[8] The values for r_{H_2-p}, r_{H_2-c} and r_{H_2-ng} can be taken depending on the utilization ratios of the petroleum, coal, and natural gas presented in the literature.

In order to estimate the level of global peace quantitatively, it is assumed that the highest value of global peace is equal to 1, which is a reference point to evaluate the interactions between global unrest and global peace.[12] Thus, the relationship between global unrest and global peace can be written as a function of the utilization ratio of hydrogen from nonfossil fuels. The general algebraic case form of global peace expression proposed is as follows:

$$GP = \frac{U_{H_2}}{Q_{wpec}} \times \frac{1}{(GU)} \quad (6)$$

where U_{H_2} defines the utilization of hydrogen from nonfossil fuels, which can be taken from the literature.

CASE STUDY

The first case study is given to determine the green-energy-based sustainability ratio depending on the three subcases that are carried out based on the sectoral impact ratio, the technological impact ratio, and the practical application impact ratio by using actual data from the literature.[8] All cases are presented in Table 1. The results of the first case study are presented in Figs. 7 and 8a–c.

The second case study is performed to determine the global unrest and global peace level based on the predicted utilization ratios of hydrogen from nonfossil fuels. In this regard, two important empirical relations that describe the effects of fossil fuels on world peace and global unrest are taken into consideration for this case study. The results of this case study are presented in Fig. 9.

Results and Discussion

In accordance with the objective of this article, the projection data is obtained using actual data of primary energy, fossil fuel, and green energy consumptions. Considering the technological impact ratio (max. value = 1/3 of the green energy financial budget), the sectoral impact ratio (max. value = 1/3 of the green energy financial budget allocation), and the practical application impact ratio (max. value = 1/3 of the green energy financial budget allocation), three cases are analyzed and discussed in detail in this part.

Fig. 7 shows the variations of the fossil fuel utilization ratio (world fossil fuel consumption/world primary energy consumption) as a function of the green energy utilization ratio (world green energy consumption/world primary energy consumption). It is found out from this figure that the fossil fuel utilization ratio decreases depending on the rise of the year while green energy utilization ratio increases. For example, the green energy utilization ratio was 5.71% in 1970, 8.25% in 1980, 11.67% in 1990, and 13.27% in 2000 while the fossil fuel utilization ratio was 94.29% in 1970, 91.75% in 1980, 88.32% in 1990, and 86.77% in 2000, based upon the actual data. However, it is observed that the green energy utilization ratio increased and reached 12.31% while fossil fuel utilization ratio decreased to 87.69% in 2004. Based on the projected data, it expected that green energy utilization ratio will reach

Table 1 The cases depending on green energy impact ratio for the green-energy-based sustainability ratio

Effect of variable parameters (ε)			10%	30%	50%	70%	90%
Cases		Percent of total financial budget for green energy	Green energy impact ratio (R_{gei})				
I	2 constant parameters	$n/3$	0.699	0.766	0.833	0.899	0.966
	1 variable parameter	$k\times(1/3)\times\varepsilon$					
II	1 constant parameter	$n/3$	0.400	0.533	0.666	0.800	0.933
	2 variable parameters	$k\times(1/3)\times\varepsilon$					
III	3 variable parameters	$k\times(1/3)\times\varepsilon$	0.100	0.300	0.500	0.700	0.900

Here, n defines the number of constant parameters; k, the number of variable parameters.

almost 13.52% in 2006, 14.09% in 2012, 14.58% in 2018, 15.01% in 2024, and 15.38% in 2030. However, it is expected that the fossil fuel utilization ratio will decrease to almost 86.48% in 2006, 85.91% in 2012, 85.42% in 2018, 85.99% in 2024, and 84.62% in 2030. Thus, in order to increase the green energy utilization ratio and to reduce the harmful effects resulting from the fossil fuel consumption, the investments on green energy should be encouraged and the green energy strategies should be put into practice for sustainable development.

First, it should be stated that one or two of the parameters that are sectoral, technological, and practical application impact ratios can be selected as constant parameters. As shown in Table 1, a variable and two constant parameters are considered in Case 1, two variables and one constant parameter are considered in Case 2, and three variable parameters are considered in Case 3. When Case 1 is applied for green energy supply and progress, it is found that the green energy impact ratio changes between 0.699 and 0.966 depending on the percentages of the variable parameter. In the application of Case 2 it is obtained that the green energy impact ratio varies between 0.40 and 0.933 depending on the percentages of two variable parameters. In the application of Case 3, it is calculated that the green energy impact ratio changes from 0.1 to 0.9 depending on the percentages of three variable parameters. When the three Cases are compared to each other, it can be said that the highest values of green energy impact ratio are found by applying Case 1, and also that Case 3 gives the lowest green energy impact ratios. Thus, Case 1 should be selected to

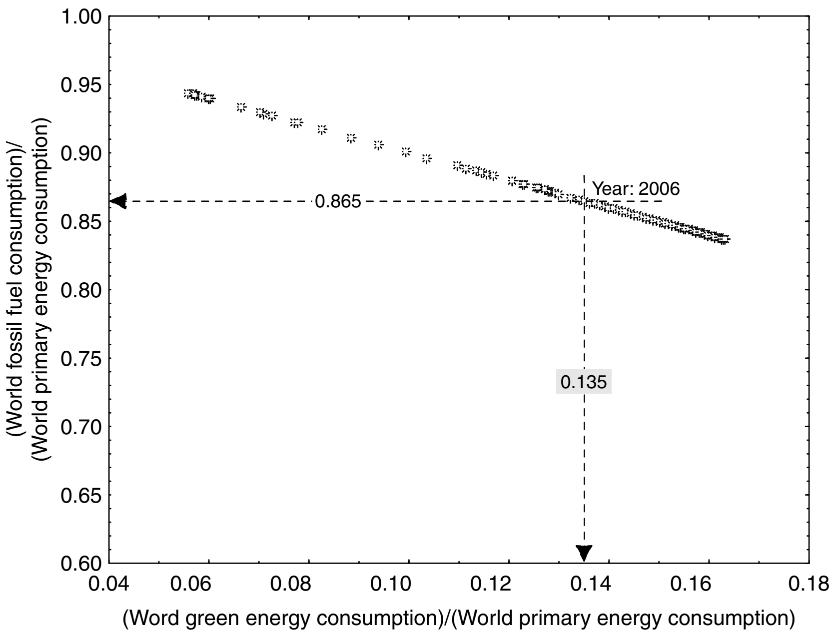


Fig. 7 World fossil fuel consumption ratio as a function of green energy consumption ratio.

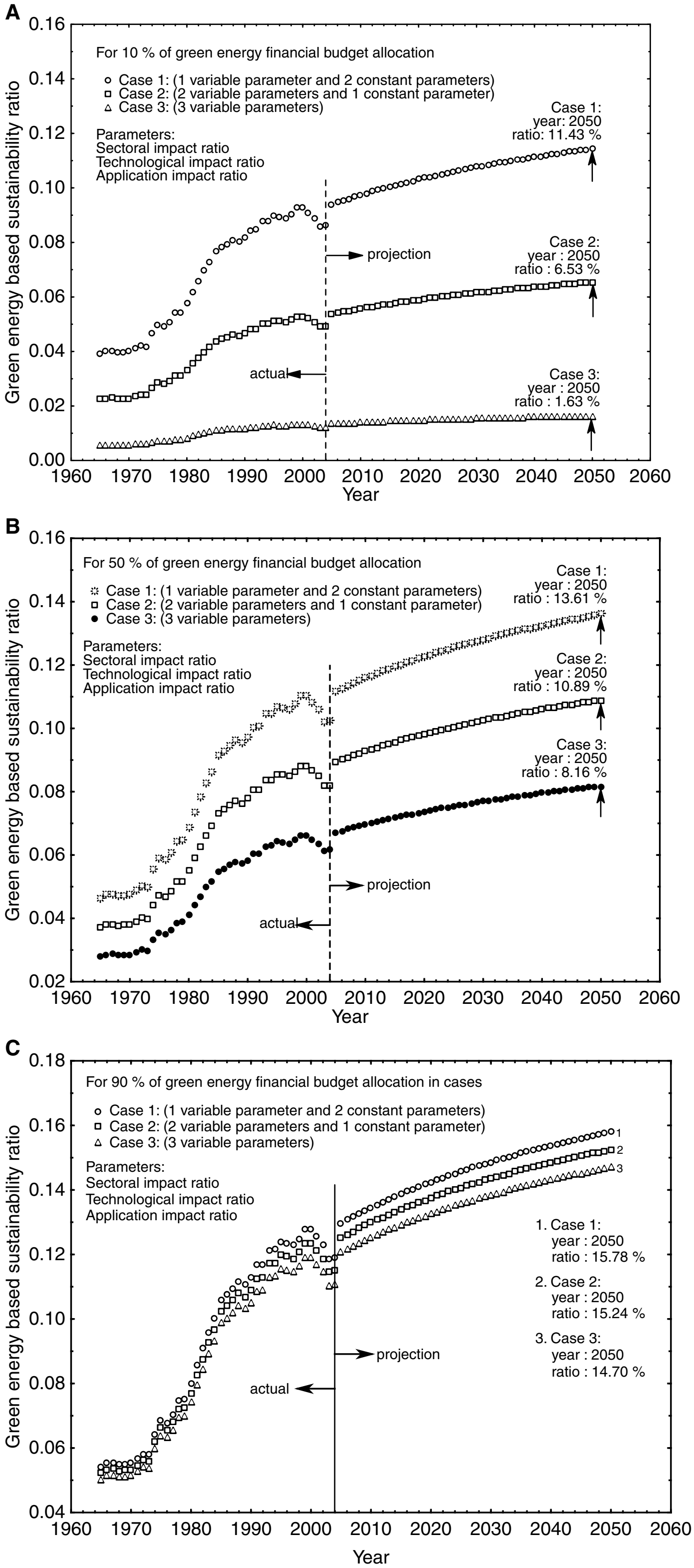


Fig. 8 (a) Green-energy-based sustainability ratio versus time depending on the actual and projected green energy consumption data for 10% of green energy financial budget allocation. (b) Green-energy-based sustainability ratio versus time depending on the actual and projected green energy consumption data for 50% of green energy financial budget allocation. (c) Green-energy-based sustainability ratio versus time depending on the actual and projected green energy consumption data for 90% of green energy financial budget allocation.

increase the green energy impact ratio and the green-energy-based sustainability ratio.

Fig. 8a–c show a variation of the green-energy-based sustainability ratio (R_{ges}) as a function of year by depending on the percentages of the green energy financial budget as 10%, 50%, and 90%, or the effect of the parameters in the Cases, respectively. The values of green-energy-based sustainability ratios were calculated using Fig. 4.

As shown in these figures, the values of R_{ges} increase with time based on the cases. The highest values of R_{ges} are

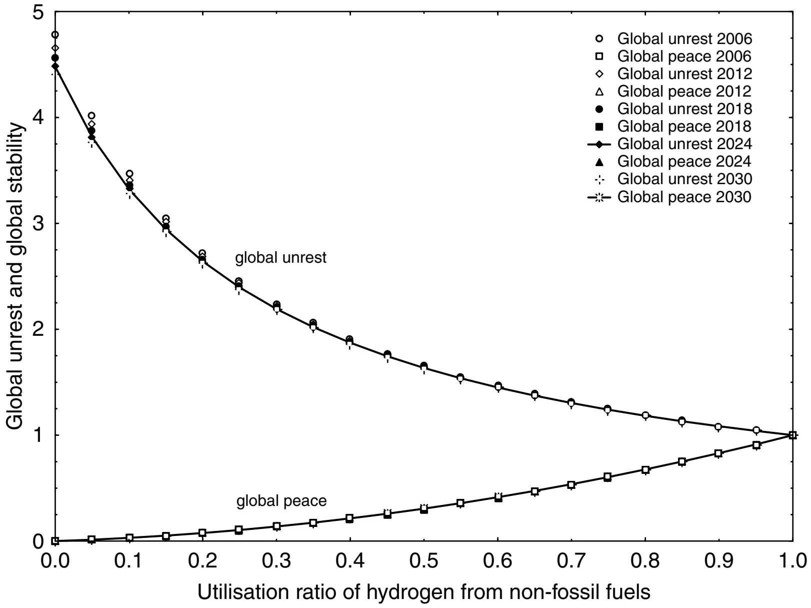


Fig. 9 Comparison of levels of global unrest and global stability as a function of hydrogen utilization ratios from nonfossil fuels.

obtained when Case 1 is applied, as shown in Fig. 8a–c. For example, the green-energy-based sustainability ratios are estimated to be 9.46% in 2006, 9.86% in 2012, 10.21% in 2018, 10.50% in 2024, and 10.76% in 2030 in case of 10% of green energy financial budget; 13.06% in 2006, 13.62% in 2012, 14.09% in 2018, 14.51% in 2024, and 14.86% in 2030 in case of 90% of green energy financial budget.

It is important to implement green energy strategies through green energy systems and applications for sustainable future. If so, the green-energy-based sustainability ratio increases and green energy is more easily supplied, thus its technologies are more preferred and applied. Hence, as long as sustainable green energy strategies are increasingly applied and the green technologies are more utilized and encouraged, the negative effects stemming from the fossil fuel utilization will decrease and the green-energy-based sustainability ratio will increase.

Fig. 9 compares the levels of global unrest and global peace as a function of the predicted utilization ratio of hydrogen from nonfossil fuels. Fig. 6 indicates that there is an inversely proportional relationship between global peace and global unrest, depending on the utilization ratio of hydrogen from nonfossil fuels.

To better appreciate this figure, some of the key energy-related reasons for global unrest need to be understood. They include:

- Increases in fossil fuel prices
- Environmental effects of energy use, including pollution due to emissions, stratospheric ozone layer depletion, and global warming
- Decreases in the amount of fossil fuel available per capita and the associated decrease in living standards
- Increases in energy demand due to technological developments attributable to and based on fossil fuels
- Depletion of fossil fuel reserves
- Increases in conflicts for fossil fuel reserves throughout the world
- The lack of affordable and practical alternative energy sources to fossil fuels

It is found out from Fig. 9 that an increase in hydrogen utilization accordingly decreases the reasons for global unrest, allowing the benefits of global peace to be realized over time; and the lowest levels of global unrest occur when hydrogen from nonfossil fuels is substituted completely for fossil fuels. In general, the level of global unrest is higher than 1 and the problems causing global unrest can be reduced by using hydrogen energy from nonfossil sources instead of fossil fuels. Fig. 5 suggests that the utilization of hydrogen from nonfossil fuels at certain ratios of petroleum, coal, and natural gas decreases the amount of fossil fuel consumption and thus reduces the level of global unrest closer to 1. As shown in Fig. 9, it is expected that, depending on the actual and projected fossil fuel consumption data, the levels of global unrest will be approximately 4.78 in 2006, 4.66 in 2012, 4.56 in 2018, 4.48 in 2024, and 4.41 in 2030 when the utilization ratio of hydrogen is zero. If the utilization ratio of hydrogen from nonfossil fuels is lower than 100%, the level of global peace is less than 1 and the reasons for global unrest increase. Therefore, it is beneficial to encourage the utilization of hydrogen from nonfossil fuels in place of fossil fuels. The highest level of global peace is attained when 100% of hydrogen from nonfossil fuels is used in place of fossil fuels. Some advantages of having the highest level of global peace are:

- Lifetimes of fossil fuel reserves are extended and real fossil fuel prices, consequently, can be held constant or reduced relative to present prices.
- Environmental effects from using fossil fuels are reduced or prevented because of the utilization of hydrogen from renewable energy sources and technologies.
- Technological developments based on hydrogen from nonfossil fuels increase and the requirement of technologies based on fossil fuels decrease.
- Living standards are probably higher than at present due in part to the increased consumption of the technologies related to hydrogen from sustainable green energy sources.
- Pressures to discover energy sources reduce because hydrogen can be abundantly produced and conflicts for energy supplies subside.

To increase global peace, the relationship of hydrogen to renewable energy sources needs to be understood, as does the importance of producing hydrogen from renewable energy sources.

Fig. 10 describes routes using sustainable green energy sources for green power production. It is expected that the utilization of sustainable green energy sources will reduce the negative energy-related environmental effects such as global climate change and emissions of CO, CO_2, NOx, SOx, nonmethane hydrocarbons, and particulate matter. In this regard, some potential solutions to decreasing the global unrest associated with the harmful pollutant emissions have evolved, including[3]:

- Education and training for a sustainable future
- Renewable energy technologies
- Energy conservation (efficient energy utilization)
- Cogeneration and district heating
- Energy storage technologies
- Alternative energy dimensions for transport
- Energy source switching from fossil fuels to environmentally benign energy forms
- Coal cleaning technologies
- Optimum monitoring and evaluation of energy indicators
- Policy integration
- Recycling
- Process change and sectoral shiftment
- Acceleration of forestation
- Carbon or fuel taxes
- Greener materials substitution
- Promoting public transport
- Changing life styles
- Increasing public awareness

Considering the above explanations, the following important remarks can be extracted. Fossil fuel consumption and green energy consumption are expected to reach 13807.2 and 2694.9 M, respectively, by the year 2050. This increase indicates that humans will still be dependent on fossil fuels. Based on the projected data, the green energy consumption ratio expects that the green energy utilization ratio will reach almost 16.33% and the fossil fuel utilization ratio will decrease to almost 83.67% in 2050. If the increase of fossil fuel consumption continues in this manner, it is likely that the world will be affected by many negative problems due to fossil fuels. More utilization of fossil fuels will harm world stability and increase local and global environmental problems, resulting in increasing global unrest. It is thus suggested

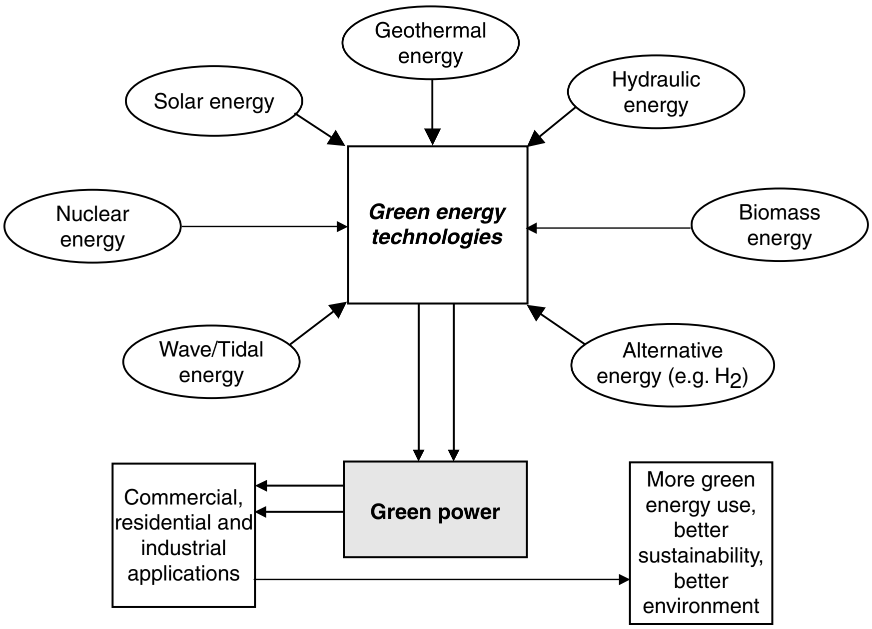


Fig. 10 Routes for green power production from sustainable green energy sources.

that the utilization of fossil fuels should be reduced, and fossil-based technologies should be gradually converted to green-energy-based technologies.

Case 1 gives better results than Case 2 and Case 3. Therefore, for a higher green energy impact ratio in practice, Case 1 should be applied to increase the green energy sustainability ratio depending on the green energy strategies. Moreover, Case 1 gives the best results of the green-energy-based sustainability ratio depending on the green energy impact ratio and the green energy utilization ratio.

The approximate quantified measures developed for level of global peace (ranging between 0 and 1) and level of global unrest (ranging between 1 and ∞) expressions can help understand and measure levels of global unrest and global peace. The highest level of global peace occurs when GP $=1$ and correspondingly, the lowest level of global unrest when GU $=1$, and efforts to increase global peace and stability should cause the values of GP and GU to shift towards these limiting cases. Hydrogen from nonfossil fuels can replace oil, coal, and natural gas to reduce the level of global unrest.

Sustainable green energy strategies are definitely required to ensure the global stability by reducing the harmful effects of fossil-based energy consumption. So, it is suggested that the importance of green energy and technologies that probably reduce world problems and achieve a sustainable energy system should be emphasized with consideration of the sustainable energy strategies. Moreover, a transition to the green-energy-based economy should be encouraged and developed countries in particular should increase investments in green energy and technologies. Progress of green energy and technologies is based on sustainable green energy strategies for future green energy scenarios. The foremost factor that will determine the specific role of green energy and technologies will likely be energy demand. In order to balance the energy demand now and in the future, it is suggested thatsustainable green energy sources and technologies be taken into consideration to increase the sustainable development in a country.

CONCLUSION

Green energy for sustainable development has been discussed and some key parameters to increase green-energy-based sustainability and the global peace level have been presented. The effects of technological, sectoral, and practical application impact ratios on the green energy impact ratio and the green-energy-based sustainability ratio are examined thoroughly. In addition, the key role of hydrogen—one of the most green energy carriers for the future—is discussed in terms of global unrest and peace. Accordingly, sustainable green energy and technologies are definitely required to ensure global stability by reducing the harmful effects of fossil-based energy consumption. The most important scenario that encourages the transition to green energy and technologies and promotes green-energy-based technologies is to supply the required incentives and interactions among the countries, scientists, researchers, societies, and all. Therefore, the investments in green energy supply should be, for the future of world nations, encouraged by governments and other authoritative bodies who, for strategic reasons, wish to have a green alternative to fossil fuels.

REFERENCES

1. Midilli, A.; Ay, M.; Dincer, I.; Rosen, M.A. On hydrogen and hydrogen energy strategies: I: current status and needs. Renewable Sustainable Energy Rev. **2005**, *9* (3), 255–271.
2. Midilli, A.; Dincer, I.; Ay, M. Green energy strategies for sustainable development. Energy Policy **2006**, *34* (18), 623–633.
3. Dincer, I. Renewable energy and sustainable development: a crucial review. Renewable Sustainable Energy Rev. **2000**, *4* (2), 157–175.
4. Dincer, I. The role of exergy in energy policy making. Energy Policy **2002**, *30*, 137–149.
5. Dincer, I.; Rosen, A.M. Exergy as a driver for Achieving Sustainability. Int. J. Green Energy **2004**, *1* (1), 1–19.
6. Midilli, A.; Ay, I.; Dincer, M.; Rosen, M. A.; *On Hydrogen Energy Strategy: The Key Role in This Century and Beyond*, Proceedings of The First Cappadocia International Mechanical Engineering Symposium (CMES-04), *1*, 211–217, Cappadocia, Turkey, 2004.
7. World Energy Outlook, EO&W, Institute of Shipping Analysis, http://saiweb.sai.se, Available at http://www.worldenergyoutlook.org/weo/pubs/weo2002/WEO21.pdf, 2002.
8. Workbook, Statistical Review Full Report of World Energy 2005, Available at http://www.bp.com/centres/energy, 2005.
9. Dincer, I. Environmental impacts of energy. Energy Policy **1999**, *27*, 845–854.
10. Dincer, I.; Rosen, A.M. Thermodynamic aspects of renewables and sustainable development. Renewable Sustainable Energy Rev. **2005**, *9*, 169–189.
11. Rosen, M.A.; Scott, D.S. Entropy production and exergy destruction—part I: hierarchy of earth's major constituencies, and part II: illustrative technologies. Int. J. Hydrogen Energy **2003**, *28* (12), 1307–1323.
12. Midilli, A.; Ay, M.; Dincer, I.; Rosen, M.A. On hydrogen and hydrogen energy strategies II: future projections affecting global stability and unrest. Renewable Sustainable Energy Rev. **2005**, *9* (3), 273–287.

Greenhouse Gas Emissions: Gasoline, Hybrid-Electric, and Hydrogen-Fueled Vehicles☆

Robert E. Uhrig
University of Tennessee, Knoxville, Tennessee, U.S.A.

Abstract
In another chapter of this encyclopedia (Hybrid-Electric Vehicles: Plug-In Configuration), the author showed that it was theoretically possible for plug-in hybrid-electric light transportation vehicles to utilize electricity provided by electric utilities to displace almost 75% of the energy of gasoline used by light transportation vehicles. (Uhrig, R.E. Using plug-in hybrid vehicles to drastically reduce petroleum-based fuel consumption and emissions. The BENT of Tau Beta Pi, 2005; XCLI (2), 13–19.). The chapter indicated that there also is a significant reduction in the greenhouse gases emitted into the atmosphere by the plug-in hybrid vehicles, but no quantitative assessment is provided. The purpose of this chapter is to provide quantitative comparisons between all of the various hybrid-electric options. Comparisons are also made with greenhouse gas emissions from hydrogen-fueled vehicles. The results shown in tables indicate that greenhouse gas emissions are small or negligible when the electricity or hydrogen is produced using nuclear or solar (wind, photovoltaic, hydro, or renewable) energy. However, when fossil fuels (coal, oil, or natural gas) are used, the greenhouse gas emissions are greater than for gasoline-fueled vehicles in most cases.

INTRODUCTION

In another chapter of this encyclopedia (Hybrid-Electric Vehicles: Plug-In Configuration), the author showed that it was theoretically possible for plug-in hybrid-electric light transportation vehicles to utilize electricity provided by electric utilities to displace almost 75% of the energy of gasoline used for light transportation vehicles (automobiles, SUVs, pickup trucks, minivans, etc.). It was also shown that replacing this gasoline energy with electricity would require 200–250 GW of new electrical generating capacity. Calculations based on the results in another chapter in this encyclopedia (Transportation Systems: Hydrogen Fueled) show that about 930 GW of new electrical generating capacity would be required to produce hydrogen by electrolysis to replace all hydrocarbon fuels used in all U.S. transportation systems (including heavy trucks, aircraft, military vehicles, etc.). Hence, the choice of fuels for of these new electrical generating plants could have a large environmental impact if a significant fraction of gasoline used for light and heavy vehicle transportation was replaced by electricity or hydrogen.

☆ The material in this article was originally published as "Greenhouse gas emissions from gasoline, hybrid-electric, and hydrogen-fueled vehicles" by Uhrig, R.E. in the "Proceedings of the Climate Change Technology Conference," Ottawa, Ontario, Canada, May 9–12, 2006. It is presented here with permission of the Engineering Institute of Canada (EIC), the sponsoring organization.

Keywords: Greenhouse gases; Carbon dioxide; Alternate automotive fuels; Emissions.

GREENHOUSE GASES

Sunlight enters the atmosphere striking the earth, and it is reflected back towards space as infrared radiation (heat energy). Greenhouse gases absorb this infrared radiation, trapping the heat in the atmosphere. Until about 150 years ago, the amount of energy absorbed was about the same as the amount of energy radiated back into space. At that time, this equilibrium was upset, and the concentration of greenhouse gases—especially CO_2 (carbon dioxide)—began to increase. The concentration is continuing to increase at a rate that seems to parallel the increase in production of CO_2. During the same period, the average temperature of the atmosphere increased with subsequent climate changes in a manner that is thought by many to be a cause-effect relationship.[3]

The principal greenhouse gases are water vapor (60%–65%) and CO_2 (20%–25%). Water vapor stays in the atmosphere for a relatively short time—a matter of hours or days—whereas CO_2 has an average residence period of about a century. As a result, the amount of water vapor remains relatively constant related to the rate at which it is produced by weather phenomena, while CO_2 tends to accumulate as the amount emitted increases. All other greenhouse gases (10%–20%) such as methane, ozone, nitrous oxide, and carbon monoxide tend to have less effects on the atmosphere over time because of their small quantities or short residence times. Hence, the only greenhouse gas considered in this analysis is CO_2.

In contrast with the "well to wheels" approach that includes the emissions produced by processing the original fuel, this analysis starts with the fuel used to generate electricity or to produce hydrogen. This simplification does not materially change the relationship between the emissions of greenhouse gases by the various processes analyzed.

MODELS USED FOR ANALYSES

The models to quantitatively evaluate the greenhouse gas emissions, specifically carbon dioxide, when electrical energy or hydrogen is used to replace gasoline as an automotive fuel are those used in the author's previous publications.[7,8] These models were based on information provided by or extrapolated from data provided by the Department of Energy (DOE) Energy Information Administration (EIA).[2]

These models utilize the following data:

1. There were 225 million light transportation vehicles in the United States in 2004, including 133 million automobiles and 92 million light truck-based vehicles (sport utility vehicles, vans, pickup trucks, passenger minivans, and delivery vans).
2. On any given day, 50% of these vehicles traveled less than 32.2 km (20 mi).
3. The average distance traveled by each vehicle in the United States was 19,749 km/year (12,264 mi/year). This distance traveled by each average vehicle is derived from the 9 million barrels of oil per day used to make gasoline for the 225 million light vehicles having an average fuel consumption of 8.51 km/l (20 mi/gal), the approximate average gasoline mileage reported by EIA.[2]
4. The quantitative index used for comparing electricity and hydrogen produced by different methods using different source fuels is kilograms of CO_2 per vehicle-year (pounds per vehicle-year) for the average distance traveled by each vehicle in the United States—19,749 km/year (12,264 mi/year). Actual emissions of specific vehicles may be greater or less than the values presented here depending upon whether the fuel consumption of the particular vehicle is greater or less than the average value of 8.51 km/l (20 mi/gal).

CO_2 EMISSIONS FOR GASOLINE-FUELED LIGHT VEHICLES

TerraPass, a vendor dealing in carbon dioxide emission credits, has developed a "carbon dioxide calculator" that gives the emission of carbon dioxide for virtually all American automobiles for the past 20 years.[6] Experimenting with this calculator shows that the quantity of carbon dioxide, the most important greenhouse gas from combustion of gasoline in internal combustion engines, emitted per unit of time is a direct function of the amount of gasoline used in that time. This calculator utilized the fact that the combustion of 1 gal of gasoline emits 2.35 kg of CO_2 per liter (19.56 lbs of CO_2 per gallon) of gasoline. Because the average gas mileage in our model is 8.51 km/l (20 mi/gal), the annual emission of carbon dioxide is

$$\frac{19,749\ \text{km/year} \times 2.35\ \text{kg}\ CO_2\ \text{per liter}}{8.51\ \text{km/liter}}$$
$$= 5,458\ \text{kg/vehicle-year}$$
$$(11,997\ \text{lb}\ CO_2\ \text{per vehicle-year})$$

Clearly, the amount of CO_2 for smaller vehicles is less than for larger vehicles because the gasoline mileage is greater. However, the model for this analysis deals with average vehicles that achieve 8.51 km/l (20 mi/gal)—the reference performance used for all comparisons.

CO_2 EMISSIONS FOR MICRO, MILD, AND FULL HYBRID VEHICLES

Because of the energy to propel all traditional types of hybrid vehicles is provided by gasoline, we will utilize the increased fuel mileages assigned in the author's earlier publication[8] and reduce the total CO_2 emissions for traditional hybrid-electric vehicles accordingly, as shown in Table 1.

These values are for hybrid vehicles of a size corresponding to a traditional vehicle that attains 8.51 km/l (20 mi/gal). A larger or smaller vehicle would have higher or lower emissions, respectively.

CO_2 EMISSIONS FOR PLUG-IN HYBRID-ELECTRIC VEHICLES

The model utilized in the author's hybrid article[8] has half the vehicles traveling 24.2 km (15 mi) per day on electricity, or 8816 km (5475 mi) per year. The other half of the vehicles travel 56.4 km (35 mi) per day, or 20,572 km (12,775 mi) per year on electricity. These assumptions result in an average of 14,694 km (9125 mi) per year on electricity for all vehicles.

Table 1 Emissions for hybrid-electric vehicles

	Gasoline mileage		CO_2 per vehicle-year	
Type hybrid	**km/l**	**mi/gal**	**Kg**	**lb**
Micro hybrid	9.35	22	4,955	10,906
Mild hybrid	10.62	25	4,362	9,598
Full hybrid	12.32	29	3,761	8,274

The second half of these vehicles must also travel an additional 5055 km (3139 mi) per year as a full hybrid using gasoline at 12.32 km/l (29 mi/gal). Because the average total distance traveled by each car per year is 19,749 km (12,264 mi), the distance traveled using gasoline is 25.6% of the average distance each vehicle travels. The other 74.4% of this distance is traveled using electricity generated by a utility using nuclear fuels, fossil fuel, solar energy (wind, hydro, or photovoltaic systems), or renewable energy systems. Such a substitution of electricity for gasoline would save 6.7 million of the 9.0 million barrels of oil per day used in the United States for light vehicle transportation.

Electricity Generated Using Nuclear, Solar, and Renewable Fuels

Plug-in hybrid vehicles have two sources of CO_2—the CO_2 emitted due to operation in the full-hybrid mode and the CO_2 emitted in generating the electrical energy used for the rest of the time. Because nuclear, solar, and renewable fuels generate a net of zero CO_2, the only CO_2 generated is by operation in the full-hybrid mode for 25.6% of the distance traveled. Hence, the emission is 25.6% of the value in Table 1 for a full-hybrid, 964 kg (2122 lb) per vehicle-year.

Electricity Generated Using Fossil Fuels (Natural Gas, Oil, and Coal)

When the electricity is supplied by utilities using fuels that emit carbon dioxide during combustion, this CO_2 from generating electricity must be added to the CO_2 of the full-hybrid operation discussed above. Data on actual emissions from fossil power plants in the United States for 1999 (the last year for which complete data is available), provided by EPA and (DOE) (DOE–EIA 1999), show the emission rates in Table 2 below.

In a previous Chapter (Hybrid-Electric Vehicles: Plug-In Configuration), it was established that 0.374 kWh/km (0.603 kWh/mi) was a reasonable average expenditure of electrical energy for hybrid vehicles that would be comparable to a vehicle getting 8.51 km/l (20 mi/gal) of gasoline using an internal combustion engine. Previously in this chapter, it was shown that the average distance traveled by all plug-in hybrid vehicles while operating on electricity alone was 14,694 km/year (9125 mi/year). Hence, the electricity used is

$$(14,694\ \text{km/year}) \times (0.374\ \text{kWh/km}) = 5,496\ \text{kWh/year}.$$

If we multiply this value by the amount of CO_2 emission per kilowatt-hour for the three fossil fuels given in Table 2 and add the 964 kg/vehicle-year (2122 lb/vehicle-year) of CO_2 for operation in the hybrid mode, we get the results for fossil fuels given in Table 3.

Table 2 Emission rates for fossil fuels

Fuel to generate electricity	% of total generation	CO_2 emission rate	
		kg/kwh	lb/kwh
Coal	51.0	0.952	2.095
Oil	3.2	0.895	1.969
Natural gas	15.2	0.600	1.321
Renewables	0.6	0 (net)	0 (net)
Non-fossil[a]	30.0	0	0

[a]Nuclear, Solar, Wind, and Hydro.
Source: From Energy Information Administration (see Ref. 1).

HYDROGEN FUELED VEHICLES

The emission of carbon dioxide from vehicles utilizing electricity generated with fuel cells operating on hydrogen is negligible except for the emissions from the various processes used to produce the hydrogen. These emissions of carbon dioxide from the production processes must be taken into account to give a valid comparison with emissions of greenhouse gases from the other vehicle configurations and their potential impact upon climate and weather modification.

STEAM METHANE REFORMING

Some 95% of hydrogen used today is produced by steam methane reforming (SMR) from natural gas (∼98% methane—CH_4). The two steps of SMR are

$$CO_4 + H_2O + \text{Heat (206 kJ/mol)} \xrightarrow{\text{Catalyst}} CO + 3H_2$$

(Steam Reforming Reaction)

$$CO + H_2O \xrightarrow{\text{Catalyst}} CO_2 + H_2 + \text{Heat (41 kJ/mol)}.$$

(Water–Gas Shift Reaction)

Table 3 Emissions for fossil fuels

Fuel to generate electricity	Emissions of CO_2	
	kg/vehicle-year	lb/vehicle-year
Coal	6,205	13,650
Oil	5,889	12,956
Natural gas	4,271	9,396

Source: From Energy Information Administration (see Ref. 1).

The hydrogen comes from the methane and the steam. The steam reforming reaction is endothermic with the required 206 kJ/mol heat energy normally produced by combustion of some of the methane. The water–gas shift reaction is exothermic, providing 41 kJ/mol heat energy that, if recovered, can reduce the amount of methane burned. Then the methane burned has to provide only the net 165 kJ/mol that represents ~17% of the total methane energy. About 3.3 mol of hydrogen (~83% of the theoretical maximum of 4 mol of hydrogen from CH_4) are produced for each mol of methane. A well-designed SMR plant will yield hydrogen having about 80% of the energy of the methane supplied. Unfortunately, SMR produces CO_2 in both the methane combustion and in the water–gas shift reaction.

Emissions of Vehicles Using Hydrogen Produced by SMR

In an earlier publication,[8] it was established that 0.228 kWh/km (0.367 kWh/mi) of mechanical energy at the tire-pavement interface corresponded to 8.51 km/l (20 mi/gal). If we use the following efficiencies:

70% efficiency for the electric motor drive (electricity to mechanical energy at the tire-pavement interface), the same used in the above reference,
60% efficiency (hydrogen to electricity) for the fuel cell, and
85% efficiency is distributing and dispensing the hydrogen to the fuel cell, the needed energy of the hydrogen input to the fuel cell is

$$\frac{0.228 \text{ kWh/km}}{0.70 \times 0.60 \times 0.85} = 0.638 \text{ kWh/km } (1.028 \text{ kWh/mi})$$

of hydrogen energy. Because the average total distance traveled per vehicle is 19,749 km/year (12,264 mi/year), the total kilowatt-hour per vehicle-year is

$$(19,749 \text{ km/year}) \times (0.638 \text{ kWh of } H_2 \text{ energy/km}) = 12,607 \text{ kWh of } H_2 \text{ energy/vehicleyear.}$$

The energy of hydrogen (lower heating value) is 119.9 MJ/kg (51,600 Btu/lb), so the amount of H_2 used per year is

$$\frac{12,607 \text{ kWh per vehicle-year}}{119.9 \text{ MJ/kg} \times 0.278 \text{ kWh/MJ}} = 379 \text{ kg } H_2 \text{ per vehicle-year } (834 \text{ lb } H_2 \text{ per vehicle-year}).$$

Goswami indicates that SMR produces a net of 0.43 mol of CO_2 for each mol of H_2 produced using SMR.[4] Because the molecular weights of CO_2 and H_2 are 44 and 2, respectively, the specific CO_2 emission can be calculated by

$$(0.43 \text{ mol } CO_2/\text{mol } H_2) \times \frac{(44 \text{ gm/mol } CO_2)}{(2 \text{ gm/mol } H_2)} = 9.46 \text{ gm } CO_2/\text{gm } H_2 = 9.46 \text{ kg } CO_2/\text{kg } H_2$$

and the total emission of CO_2 per vehicle-year is

$$(379 \text{ kg } H_2 \text{ per vehicle-year}) \times (9.46 \text{ kg } CO_2 \text{ per kg } H_2) = 3,585 \text{ kg } CO_2 \text{ per vehicle-year } (7,888 \text{ lb } CO_2 \text{ per vehicle-year}).$$

Emission for Vehicles Using Hydrogen Produced by SMR with Heat Provided by a High-Temperature Nuclear Reactor

Recent work in Japan has demonstrated the feasibility of substituting high-temperature heat from a gas cooled nuclear reactor to replace the heat supplied by the combustion of methane. This increases the amount of hydrogen produced to 4 moles per mole of methane and eliminates the CO_2 produced by combustion of methane, but not the CO_2 produced by the water–gas shift reaction. The overall reaction of the steam reforming and water shift reactions is the production of 1 mole of CO_2 and 4 mol of hydrogen for each mole of CH_4. Hence, multiplying this ratio by the molecular weights and the amount of hydrogen used per year gives

$$\frac{1 \text{ mol } CO_2}{4 \text{ mol } H_2} \times \frac{44}{2} \times 379 \text{ kg } H_2 \text{ per vehicle-year} = 2,084 \text{ kg } CO_2 \text{ per vehicle-year } (4,586 \text{ lb } CO_2 \text{ per vehicle-year}).$$

EMISSIONS FOR VEHICLES USING H_2 PRODUCED BY ELECTROLYSIS

Conventional electrolysis of water to produce hydrogen is a well-developed technology and production units as large as 10 MWe are commercially available today. However, the typical overall efficiency of hydrogen production using electrolysis based on the thermal content of the generating plant fuel is about 25% today, consisting of two components—about 33% efficiency in converting fossil or nuclear fuel to electricity and about 75% in using electricity to separate water into hydrogen and oxygen. If a high-temperature gas-cooled reactor or a modern high-efficiency gas-fired combined cycle plant is used, the overall efficiency in producing hydrogen could approach 45%.

A leading manufacturer of electrolysis equipment indicated that 1 MW of electricity can generate 0.52 ton (1040 lb) of hydrogen per day or 0.473 kg H_2/kW day (1.04 lb H_2/kW day).[5] In the case of SMR, it was calculated that the hydrogen required per year for a vehicle using hydrogen to travel 12,260 mi/year was 379 kg of H_2/vehicle-year (832 lb/vehicle-year). Hence, the amount of electricity required per year is

$$\frac{(379 \text{ kg } H_2/\text{year}) \times (24 \text{ h/day})}{0.473 \text{ kg } H_2/\text{kW day}} = 19,230 \text{ kWh/year}.$$

The carbon dioxide emitted for hydrogen-fueled vehicles is the carbon dioxide emitted in producing the electricity required to produce the hydrogen. The choices of fuel for generating electricity are coal, oil, natural gas, nuclear energy, and solar energy (wind, photovoltaics, and hydro). Hence, the emissions of carbon dioxide are the product of the kilowatt-hour per year and the appropriate emission per kilowatt-hour from Table 2 for the fuel used.

Emissions for Vehicles Using H_2 Produced by Electrolysis Using Solar, Nuclear, and Renewable Energy

Because none of these energy sources emit carbon dioxide when generating electricity, there are no emissions of greenhouse gases associated with these arrangements.

Emissions for Vehicles Using H_2 Produced by Electrolysis Using Fossil Energy (Coal, Oil, and Natural Gas)

The emissions of carbon dioxide for these fossil fuels are the products of the 19,230 kWh/year and the appropriate emission per kilowatt-hour for the fuel as given in Table 2. The results are shown in Table 4.

Table 4 Emissions for fossil electrolysis

Fuel	kg CO_2/year	lb CO_2/year
Natural gas	11,538	25,384
Oil	17,211	37,963
Coal	18,307	40,276

EMISSIONS FOR VEHICLES USING H_2 PRODUCED BY THERMOCHEMICAL PROCESSES

The overall efficiency of the thermochemical process of producing hydrogen, such as the sulfur–iodine process, approaches 50% at 900°C, while the overall efficiency of the electrolysis process is about 25%. Hence, for the same fuel, the carbon dioxide emissions will be half as much for thermochemical processes as for electrolysis.

Emissions for Vehicles Using Hydrogen Produced by Thermochemical Processes Using Solar, Nuclear, and Renewable Energy

Because none of these energy sources emit carbon dioxide when generating electricity, there are no emissions of carbon dioxide associated with these arrangements.

Emissions for Vehicles Using Hydrogen Produced by Thermochemical Processes Using Fossil Energy (Natural Gas, Oil, and Coal)

The emissions of carbon dioxide for these fossil fuels are half of those for electrolysis, as shown in Table 4. The results are shown in Table 5.

RESULTS AND CONCLUSIONS

The results of this analysis are presented in Tables 6 and 7, which provide emissions of carbon dioxide for the various hybrid and hydrogen-fueled vehicle arrangements for the reference case of 8.50 km/l (20 mi/gal) for the average vehicle being driven 19,749 km/year (12,264 mi/year).

Plug-in Hybrid-Electric Vehicles

In Table 6, the vehicle emissions of CO_2 for all traditional hybrid-electric vehicles are inversely related to the gas mileage. The gasoline mileage for micro, mild, and full hybrids are assumed to be 9.35, 10.62, and 12.32 km/l (22, 25, and 29 mi/gal), respectively. These are reasonable overall values for the newer, larger hybrid vehicles that have larger gasoline engines and electric motors such as the Toyota Highlander and Honda Accord hybrids.

Table 5 Emissions for hydrogen produced by thermochemical processes by fossil fuels

Fuel	kg CO_2/year	lb CO_2/year
Natural gas	5,769	12,692
Oil	8,605	18,932
Coal	9,153	20,138

Table 6 Comparison of emissions of carbon dioxide for hybrid vehicles

Vehicle emissions of carbon dioxide	kg CO_2/vehicle-year	lb CO_2/vehicle-year
Reference gasoline vehicle average 8.51 km/l (20 mi/gal)	5,500	12,000
Micro-hybrid vehicle 9.35 km/l (22 mi/gal)	4,950	10,900
Mild-hybrid vehicle 10.62 km/l (25 mi/gal)	4,350	9,600
Full-hybrid vehicle 12.32 km/l (29 mi/gal)	3,750	8,300
Plug-in hybrid vehicle		
Electricity generated with nuclear	950	2,100
Electricity generated with solar energy	950	2,100
Electricity generated with renewable	950	2,100
Electricity generated with natural gas	4,250	9,400
Electricity generated with oil	5,900	13,000
Electricity generated with coal	6,200	13,700

Vehicle travels 19,749 km (12,264 mi) per year.

The rounded emissions in Table 6 for electricity generated using oil and coal are greater than for the reference case of gasoline. However, the emission of a plug-in hybrid vehicle where the electricity is generated by nuclear or solar energy is less than 20% of the reference gasoline case. Hence, the only way to significantly improve the greenhouse gas situation if plug-in hybrids are implemented on a large scale is to use only nuclear energy, solar (wind, hydro, or photovoltaic) energy, or renewables to generate the needed electricity. Because hydro power is virtually out of the question due to limited sites and environmental concerns, renewables have large land requirements, and both photovoltaic and wind are intermittent in nature with availabilities of less than about 30%, nuclear power would appear to be the primary choice.

Hydrogen-Fueled Fuel Cell Vehicles

The vehicle emissions of CO_2 for hydrogen-fueled fuel-cell-driven vehicles are presented in Table 7, where the emissions have been rounded off. Because fuel cells do not

Table 7 Comparison of emissions of carbon dioxide for hydrogen fueled fuel cell vehicles

Vehicle emissions of carbon dioxide	kg CO_2/vehicle-year	lb CO_2/vehicle-year
Reference gasoline vehicle utilized an average 8.50 km/l (20 mi/gal)	5,500	12,100
Vehicle uses hydrogen produced using SMR	3,600	7,900
Vehicle using hydrogen produced using SMR with heat supplied by		
High temperature gas cooled reactor	2,100	4,600
Vehicle using hydrogen produced using electrolysis with electricity		
Produced using natural gas	11,550	25,400
Produced using oil	17,200	37,900
Produced using coal	18,300	40,300
Produced using nuclear energy	0	0
Produced using solar energy (wind, hydro, and photovoltaic)	0	0
Produced using renewables	0	0
Vehicle using hydrogen generated using a thermochemical process with heat		
Generated using natural gas	5,750	12,700
Generated using oil	8,600	18,900
Generated using coal	9,150	20,100
Generated using nuclear energy	0	0
Generated using solar energy (wind, hydro, and photovoltaic)	0	0
Generated using renewables	0	0

Vehicle travels 19,749 km (12,264 mi) per year.

emit CO_2, the only emissions are from the processes used to produce the hydrogen. The carbon dioxide emissions for hydrogen produced by steam methane reforming, electrolysis, and thermochemical methodologies are given.

Perhaps the most important observation is that, contrary to the widespread belief that hydrogen-fueled vehicles produce no greenhouse gases, the emissions for hydrogen-fueled vehicles vary widely depending upon the method used to generate the electricity or heat to produce the hydrogen. In the case of electrolysis, in which the electricity is generated using nuclear, solar, and renewable energy, the average annual emissions per vehicle-year are negligible. However, the emissions for fossil fuels (natural gas, oil, and coal) are about 11,500, 17,200, and 18,300 kg/year (25,300, 37,800, and 40,200 lb/year), respectively—almost three times those for plug-in hybrid vehicles. These are extremely large emissions compared to almost any other method investigated here.

The emissions for producing hydrogen for a hydrogen-fueled vehicles using SMR produces CO_2 by two methods: combustion of some of the methane to drive the reaction and the water–gas shift reaction. These reactions produce about 3600 kg (7900 lb) CO_2 per vehicle-year, about the same as a full-hybrid vehicle.

If the heat required to drive the reaction in steam methane reforming can be supplied by an outside source such as a high temperature gas cooled reactor, then the emission of CO_2 is reduced to about 2100 kg (4600 lb) per vehicle-year.

The conclusion is that using hydrogen in fuel cells to propel vehicles is a complex process with many steps (converting thermal energy to electricity to hydrogen to electricity to propelling the vehicle), each consuming considerable energy. Hence, the resulting CO_2 emissions when fossil fuels are used to generate the electricity for electrolysis or thermochemical processes are correspondingly high. Given the concerns about the influence of CO_2 on weather and climate, it seems compelling that fossil fuels not be used in producing electricity for electrolysis or thermochemical processes to produce hydrogen for transportation.

REFERENCES

1. DOE–EIA, Department of Energy, Energy Information Administration (DOE and EPA), *Carbon Dioxide Emissions from the Generation of Electric Power in the United States*, July 2000.
2. DOE–EIA, Energy Information Administration, Petroleum Flow 2002. *Annual Energy Review*, 2003; at www.eia.doe.gov/emeulaer/diagram2.html (accessed January 2006).
3. DOE–EIA, Energy Information Administration, *Greenhouse Gases, Climate Change, and Energy*, 2005; at http://www.eia.doe.gov/oiaf/1605/ggccebro/chapter1.html (accessed May 2006).
4. Goswami, D.Y. *Hydrogen Supply Technologies: Introduction to Issues*, 2005; www.ases.org/hydrogen_forum03/Goswami.pdf (accessed June 2005).
5. Stuart, S. *Nuclear-Electric Synergies: For Neat Hydrogen Production and Cleaner Fossil Reserves Harvesting*; ANS President's Special Session: Hydrogen Systems, ANS Meeting: Reno, NV, 2001.
6. TerraPass, *Carbon Dioxide Calculator*, Terrapass.com/carboncalc.html (accessed January 2005).
7. Uhrig, R.E. Engineering challenges of the hydrogen economy. The BENT of Tau Beta Pi, 2004; XCL (2), 10–19.
8. Uhrig, R.E. Using plug-in hybrid vehicles to drastically reduce petroleum-based fuel consumption and emissions. The BENT of Tau Beta Pi, 2005; XCLI (2), 13–19.

Heat and Energy Wheels

Carey J. Simonson
Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Abstract
This article discusses the construction and operation of heat and energy wheels, as well as their effectiveness in transferring heat and moisture between air streams. Heat wheels transfer sensible heat between two air streams with different temperatures, while energy wheels transfer heat and moisture between two air streams with temperature and water vapor concentration differences. Heat and energy wheels have high effectiveness values and low pressure drops, making them both economical and environmentally friendly.

INTRODUCTION

Heat wheels, as the name implies, are rotating heat exchangers that transfer heat between two different air streams (Fig. 1). Energy wheels are very similar to heat wheels except that they are designed to transfer moisture, as well as heat, between the two air streams. They are called energy wheels (and sometimes enthalpy wheels or desiccant-coated heat wheels) because they transfer sensible energy that results from temperature differences between the two air streams and latent energy that results from water vapor concentration differences between the two air streams. Heat and energy wheels are gaining popularity in building heating, ventilating, and air conditioning (HVAC) systems mainly because of their very high energy transfer effectivenesses. Heat/energy wheels are made of many different materials for heat transfer, with desiccants utilized for moisture transfer. This article presents the construction and operation of heat/energy wheels, as well as their effectiveness in transferring heat and moisture between air streams. Economic and environmental issues are also presented.

Heat Wheels

Heat wheels are rotating exchangers that transfer sensible heat between two air streams with different temperatures. These wheels have been used for half a century in gas turbine plants and electrical power generating stations to recover thermal energy from the exhaust gases and preheat inlet combustion air, thus increasing the overall plant thermal efficiency.[1–3] Typical rotating speeds are in the order of 20–30 revolutions per minute (rpm).

Keywords: Heat transfer; Moisture transfer; Rotating exchanger; Payback; Life cycle costs; Environmental impacts; Life cycle analysis; Effectiveness.

Energy Wheels

Energy wheels transfer both heat and moisture between two air streams. They have more recently been developed for transferring heat and moisture between buildings' supply and exhaust air streams. Their market share has increased significantly in the last decade, and energy wheels make up over half of all new air-to-air heat/energy exchangers installed in buildings. Their popularity can be attributed to increases in required outdoor ventilation rates in the late 1980s and a recent emphasis on improving worker productivity and health by improving the thermal comfort and indoor air quality conditions in buildings where humidity is a key variable.[4–6]

After standard ventilation rates were sharply decreased to reduce energy use in buildings in 1975,[7] the number of buildings with air quality problems resulting from high indoor air concentrations of contaminants increased significantly. Inadequate mechanical ventilation for new air-tight buildings, constructed in the 1970s and 1980s, led to many new and unforeseen indoor air quality problems.[8] As a result, most industrial countries revised their ventilation standards to include higher outdoor ventilation air flow rates. For example, the 1989 and 2004 ASHRAE ventilation standards[9,10] specify required ventilation air flow rates, which are about three times larger than they were in 1975,[7] to maintain acceptable indoor air quality for a wide range of building types and spaces. In terms of improved productivity, it has been estimated that the annual benefit would be $55 billion if all U.S. buildings were upgraded to meet current ventilation standards.[11] The average economic payback time is expected to be 1.6 years.

Naturally, the costs and environmental impacts of energy consumption have dictated that improvements in productivity, health, comfort, and indoor air quality should be achieved with minimal energy consumption, and energy wheels have been favored because the energy associated with moisture transfer (humidification or

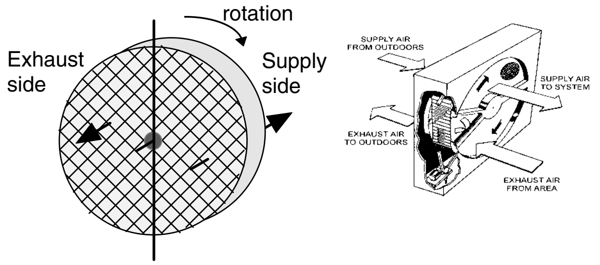


Fig. 1 Schematic of a heat/energy wheel transferring energy between the supply and exhaust air streams of a building.
Source: From 2004 ASHRAE Handbook-HVAC systems and equipment handbook. [17] American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., www.ashrae.org (see Ref. 17).

dehumidification) in building applications is often as important as heat transfer, especially in warm, moist climates. The importance of moisture transfer is evident when the cooling of moist air is considered. For example, the ideal cooling of air from 35°C and 60% relative humidity (RH) to 25°C and 50% RH requires four times as much energy as cooling air from 35 to 25°C with no change in moisture level (i.e., humidity ratio). Moisture transfer in air-to-air energy wheels can significantly reduce the dehumidification and humidification loads of buildings. This reduces the energy consumption, as well as the size of the cooling equipment needed in the building.[12]

Desiccant Drying Wheels

Desiccant drying wheels are very similar to energy wheels in that they transfer both heat and moisture between two air streams. The main difference is that desiccant drying wheels are designed with an emphasis on air-drying, rather than energy transfer. To maximize moisture transfer, desiccant drying wheels use thick desiccant coatings, slow rotational speeds (<1 rpm), and external heat sources to dry the wheel and remove as much moisture as possible from the moist air stream (supply air for buildings).[13,14] Energy wheels, on the other hand, are passive devices that transfer heat and moisture with minimal external energy input.[15]

CONSTRUCTION AND MATERIALS

Heat/energy wheels come in many different sizes and are constructed with many different materials. The core of the wheel, known as the wheel matrix, permits heat and moisture transfer between the two airstreams. One of the most common matrix materials is aluminum because of its high thermal conductivity and thermal capacitance; however, other matrix materials, such as ceramics, stainless steel, plastics, paper, and a wide range of desiccants and desiccant coatings, are also utilized. Some desiccant coatings include: zeolites, molecular sieves, silica gels, activated alumna, titanium silicate, synthetic polymers, lithium chloride, and aluminum oxide. Each matrix material and desiccant coating has a specific range of applications, performance, and limitations. For example, high temperature and corrosive applications may favor the use of stainless steel and ceramic materials, while paper and some plastics may be limited to low temperature applications.[16,17]

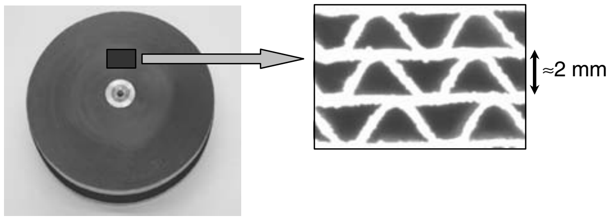


Fig. 2 A picture of an energy wheels with a typical corrugated matrix.

A common matrix arrangement is where the matrix material is corrugated to form small flow channels (e.g., sinusoidal, as in the end view of cardboard, triangular, or hexagonal) with a height of about 2 mm (Fig. 2). These small flow channels result in a large surface area for heat and mass transfer. The effectiveness of heat/energy exchangers is strongly dependent on the heat/mass transfer surface area—the larger the surface area, the greater the effectiveness. Therefore, heat/energy wheels are characterized by high heat and moisture transfer effectivenesses[15,17] because of their high heat transfer surface area to volume ratios, which result from these small flow channels. Commercial heat/energy wheels often have heat/mass transfer surface area to volume ratios of about 1000–5000 m^2/m^3. This means that for each cubic meter of exchanger, there are 1000–5000 m^2 of surface available to transfer heat and moisture. The net result is that it is possible to transfer a lot of energy with very compact wheels.

OPERATION

In a majority of installations, the two air streams flow through the heat/energy wheel in a counterflow arrangement to ensure the maximum possible heat/moisture transfer between the air streams. To explain how heat/energy wheels transfer heat and moisture between two air streams, Fig. 3 shows a side view of one of the flow channels in both a heat wheel and an energy wheel. For this example, the supply air stream is assumed to be hot and humid, while the exhaust air stream is assumed to be cool and dry. These conditions are representative of an air conditioned building during summer operating conditions.

As the hot supply air flows through the flow channel of a heat wheel, heat is transferred from the hot air to the cooler wheel matrix (Fig. 3). During this part of the wheel

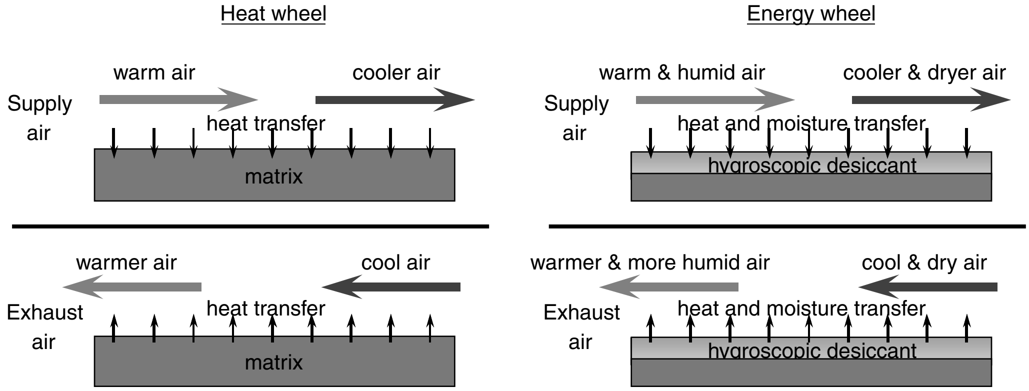


Fig. 3 A side view of a heat wheel and energy wheel matrix. Heat and moisture are transferred to the matrix from the supply air and removed from the matrix by the exhaust air.

cycle, the air stream is cooled and the wheel matrix is heated. As a result, the air leaving the wheel on the supply side is cooler than the air entering the wheel on the supply side and therefore less auxiliary energy is required to cool it for use in the building. During the other half of the wheel rotation, the cool exhaust air flows through the wheel. Here, heat is transferred from the warm matrix to the cool exhaust air. This cools the matrix and heats the air. The warm air is exhausted out of the building and the cooled matrix rotates back to the supply side, where it can cool the supply air again. This cycle repeats as the wheel continually rotates between the supply and exhaust air streams. For a typical wheel speed of 20 rpm, each flow channel in the matrix stores heat from the hot supply air and releases it to the cool exhaust air once every 3 s. Therefore, the matrix is exposed to hot air for 1.5 s followed by cold air for 1.5 s. The rotational speed of the wheel and the thermal capacity of the matrix are key parameters that affect the ability of the heat wheel to store and transfer heat between the two air streams. Other important parameters are the air flow rate and the heat transfer surface area.

The operation of an energy wheel is very similar to the operation of a heat wheel. The only difference is that the matrix of the wheel is also coated with a desiccant that stores moisture. Fig. 3 shows that moisture is transferred from the humid air stream to the hygroscopic desiccant on the supply side and removed from the desiccant by the dry air stream on the exhaust side. Here, the moisture storage capacity is an important parameter that affects the moisture transfer performance of the energy wheel. There are many types of commercial desiccants, but two of the most common are molecular sieve and silica gel. The equilibrium moisture content of silica gel is almost linearly dependent on RH and, therefore, silica gel is better-suited for applications of air-to-air energy wheels in buildings. Molecular sieve desiccants, on the other hand, are often favored in desiccant drying wheels.[18]

PERFORMANCE FACTORS

Effectiveness

Effectiveness is the most important parameter for quantifying the performance of energy exchangers, including heat/energy wheels. It is the prime factor that determines the economic viability or feasibility of an energy exchanger. Since the inlet operating conditions (temperature, humidity, and air flow rate) usually change quite slowly in typical building and other applications, the effectiveness, determined at steady-state test conditions, is used to characterize the performance of energy exchangers. The effectiveness can be measured using steady-state[19,20] or transient test methods,[21,22] but care must be taken to ensure that the uncertainty in the reported effectiveness value is low.[20,23] In most cases, it is advisable to use third-party certified wheels because the effectiveness may be verified with greater confidence and lower uncertainty.[24,25] The following equations define the three different effectiveness values for heat/energy wheels, which can range from 50 to 85% for commercial wheels.[15,17]

Sensible effectiveness for heat and energy wheels:

$$\varepsilon_s = \frac{\text{actual heat transfer}}{\text{maximum possible heat transfer}} = \frac{\dot{m}_s (T_i - T_o)|_s}{\dot{m}_{min}(T_{s,i} - T_{e,i})}. \quad (1)$$

Latent (or moisture transfer) effectiveness for energy wheels:

$$\varepsilon_l = \frac{\text{actual moisture transfer}}{\text{maximum possible moisture transfer}} = \frac{\dot{m}_s (W_i - W_o)|_s}{\dot{m}_{min}(W_{s,i} - W_{e,i})}. \quad (2)$$

Total energy transfer effectiveness for energy wheels:

$$\varepsilon_t = \frac{\text{actual energy transfer}}{\text{maximum possible energy transfer}} = \frac{\dot{m}_s(h_i - h_o)|_s}{\dot{m}_{min}(h_{s,i} - h_{e,i})} \quad (3)$$

where: ε_s is the sensible effectiveness of the heat/energy wheel; ε_l is the latent or moisture transfer effectiveness of the energy wheel; ε_t is the total effectiveness of the energy wheel; $\dot{m}$ is the mass flow rate of dry air (kg/s); T is the temperature of the air (°C or K); W is the humidity ratio of the air (kg/kg); h is the enthalpy of the air (kJ/kg); $\dot{m}_{min}$ is the minimum of the supply ($\dot{m}_s$) or exhaust ($\dot{m}_e$) air mass flow rates; and subscripts i, o, s, and e represent the inlet, outlet, supply, and exhaust sides of the heat/energy wheel.

In comparing the energy efficiency of heat/energy wheels and chillers, the recovered efficiency ratio (RER) is useful. The RER is defined in a similar manner to the energy efficiency ratio (EER) used for chillers—that is, RER equals the recovered energy rate for the exchanger at Air-Conditioning and Refrigeration Institute (ARI) test conditions[24] divided by the sum of the electrical power input for the fans to overcome the pressure drop across the wheel and the electrical power input for the motors to rotate the wheel. Values of RER range from 10 to 60 for heat wheels and from 40 to 100 for energy wheels.[15]

Pressure Drop and Leakage

All heat exchangers cause pressure losses in the fluid streams and are subject to possible leakage between the fluid streams. These factors should be considered and quantified in every design. The pressure drop across a heat/energy wheel is typically quite low compared to other air-to-air heat exchangers, with typical values ranging from 50 to 200 Pa in each air stream.[15,17] Depending on the location of the supply and exhaust fans and dampers, the pressure difference between the two air streams may be higher and result in small leakage flows and cross contamination between the supply and exhaust air streams. In addition, there is carryover of exhaust air into the supply air stream each time the wheel rotates and carries the entrained gas within the flow channels into the supply air stream. These effects can be minimized by positioning the fans so that leakage is from the fresh air on the supply side to the stale air on the exhaust side. This can be accomplished by applying good seals and installing a purge section, which uses the supply air to purge the exhaust air from the wheel before the wheel rotates into the supply air steam (Fig. 1). For some applications, toxic or hazardous contaminants in the exhaust gases will suggest that the heat/energy wheels may not be an appropriate selection, even when a purge section is used to reduce cross contamination to less than 1% of the flow rate. In other applications, large pressure differences between the exhaust and supply air ducts suggest that heat/energy wheels (as well as plate heat exchangers) will experience excessive leakage.

The ARI rating standard[24] specifies two factors that characterize the leakage between supply and exhaust air streams: the exhaust air transfer ratio (EATR) and the outside air correction factor (OACF). The EATR represents the fraction of exhaust air that is transferred (recirculated) into the supply air through the exchanger and typical values are from 1 to 10%. This means that from 1 to 10% of the exhaust air is transferred to the supply air stream and delivered to the space. The OACF represents the increase in outdoor ventilation rate required because of leakage between the supply and exhaust air streams. Typical values of OACF are from 1.0 to 1.1. An OACF of 1.05, for example, means that 5% of the supply air is transferred from the supply inlet side to the exhaust outlet side and not delivered to the building space. Therefore, the supply air flow rate must be increased by 5% to provide adequate outdoor ventilation to the building space.

Frosting

Condensation and/or frosting can occur in any air-to-air heat/energy exchanger under cold weather operating conditions. As the warm moist air passes through the exchanger, it can be cooled to the dew point temperature. Water will begin to drain from the exchanger if the temperature is above freezing, while frost will begin to accumulate within the exchanger if the temperature is below freezing. If condensation is to occur under system design conditions, a condensation drain should be provided. Excessive condensation may degrade the desiccant coating used in energy wheels. If frosting would unacceptably compromise ventilation, or if equipment would be damaged by frost, a frost control scheme should be provided. Frost control of heat/energy wheels can be achieved using several different methods, including supply air bypass or throttling, preheating the supply or exhaust air inlet flow, and wheel speed control. The manufacturer should address the issue of frost control for units intended for use in cold climates.

Energy wheels are much less susceptible to condensation and frosting than heat wheels and other sensible heat exchangers because they simultaneously transfer heat and moisture. Therefore, as the warm moist air passes through the energy wheel, it is simultaneously cooled and dried. As a result, the air is less likely to reach saturation (100% RH or dew point temperature). For many applications, outdoor temperatures as low as −30°C will not cause frosting in energy wheels, while heat wheels may experience frosting at outdoor temperatures as high as 10°C.[26,27]

Reliability

Heat wheels have a long history of very good maintenance and reliability characteristics. The experience with energy wheel applications is much shorter, which may cause some reservations about their long-term performance. Nevertheless, there are long-term experiences with desiccant drying wheels used as supply air dryers. Provided the desiccant coatings of these desiccant dryer wheels are not contaminated by solvents or excessive amounts of organics or dust, or eroded by particulates, they can have the same long-term performance characteristics as heat wheels. Although energy wheels operate at much lower temperatures than desiccant drying wheels and may use slightly different desiccant coatings, it is expected that they will have the same long-term reliability and vulnerability to airborne contaminants. The risk of such exposures will be small for well-designed systems, with correct filters being a key component.[17]

ECONOMIC AND ENVIRONMENTAL FACTORS

Economic and environmental considerations are the stimuli for the application of heat/energy wheels. This section will summarize some studies in the literature that demonstrate the payback, life cycle costs (LCC), and environmental impact of heat/energy wheels in the HVAC systems of buildings.

Payback

The time required to pay back the initial investment (payback period) depends on whether the heat/energy wheel is installed during the design and construction of a new building or the retrofit of an old building. In a new building, the size of the heating and cooling equipment can be reduced substantially when heat/energy wheels are applied. In many cases, the capital cost savings realized by downsizing the heating and cooling equipment are greater than the cost of the heat/energy wheel. This results in an immediate payback, with future energy savings realized from essentially no investment. In some cases, the cost of the new system will be greater with a heat/energy wheel than without a wheel. For these cases, payback periods of less than one year can be expected for a climate with significant heating and cooling loads, such as Chicago, Illinois.[28]

If a heat/energy wheel is to be retrofitted into a building with existing heating and cooling equipment, there is little or no possibility of reducing the size of the heating and cooling equipment unless they need replacing. In these retrofit cases, payback periods will be longer. The payback period depends strongly on the ventilation air flow rate, as well as the local climate and energy costs. Payback periods may range from two years for heat wheels to one and a half years for energy wheels in Chicago, for example.[28] Payback periods will be shorter in more severe climates (i.e., climates with more heating and cooling degree days) and will be longer in milder climates.

Life Cycle Costs

For longer life cycles (e.g., ten years for heat/energy wheels), the costs over the entire life cycle provide a more meaningful economic assessment. In fact, large savings may be lost when design decisions are based only on short-term objectives, i.e., the payback period, instead of long-term objectives, i.e., LCC.[28,29] The LCC are almost always lower for buildings with heat/energy wheels than those without heat/energy wheels. The LCC associated with supplying and conditioning outdoor ventilation air for a 120-person office building in Chicago over a ten-year life cycle are about 25%–50% lower with a heat/energy wheel than without one.[28] The energy savings over the life cycle exceed the capital costs by a factor of five. The LCC can be further reduced by about 10% by applying both a heat wheel and an energy wheel in the same system.[29] In this dual wheel system, the life cycle savings in energy exceed the capital costs by a factor of seven to eight and additional heating of the ventilation air may not be needed, even for a reasonably cold climate like Chicago, Illinois.[15]

Life Cycle Analysis

The environmental impact of many products can be assessed using the life cycle analysis (LCA) method. Life cycle analysis considers the impact of a product on the environment during its entire life cycle—from production to disposal. This includes the extraction of basic and energy raw materials, the production processes of materials and products, transportation, utilization and recycling. Life cycle analysis is similar to LCC analysis in that both address issues over the life of the product or system, rather than basing decisions on the first capital cost. However, LCA and LCC differ in their measuring metric; LCC uses money as the comparison scale, while LCA uses environmental indicators, such as carbon dioxide emissions (indicating climate change), sulfur dioxide emissions (indicating acidification potential), ethene or ethylene emissions (indicating ozone formation), and others. For many building service components, such as heat/energy wheels, the main consumption of raw materials occurs during production and the main consumption of energy occurs during use.[30–32]

Research[30–32] demonstrates that energy recovery from the exhaust air of office buildings and single-family residences is clearly an environmentally friendly solution in a cold climate (Helsinki, Finland). Energy recovery totally compensates for the harmful environmental impacts that arise from the manufacture, maintenance, and operation of the heat/energy wheel and the entire air-handling unit. A ventilation unit, with its function of providing outdoor ventilation air, but not heating the air, has a net positive impact on the environment when it includes a heat/energy wheel with an effectiveness over 20%. The greater the effectiveness, the greater the positive impact on the environment. For typical effectiveness values of 70%–75%, the emissions reduction as a result of energy recovered by the heat/energy wheel exceeds the total emissions of the air-handling unit by five to ten times. These reduced emissions result in reduced potential harmful changes to the environment. The reduction in environmental impacts due to heat/energy recovery exceeds the total impacts of the air-handling unit by two to ten times. Similar conclusions are expected in most locations, but the magnitude of the environmental benefits from heat/energy wheels will depend mainly on the local climate, energy sources, efficiencies, hours of operation, and ventilation rate. Therefore, a building with a highly effective heat/energy wheel and a higher outdoor ventilation rate may have the same environmental impact as a building with a heat/energy wheel of a lower effectiveness and lower ventilation rate. Since ventilation, indoor air quality, and health are closely related,[4] heat/energy wheels are healthy for occupants and the environment.

CONCLUSIONS

Heat wheels are rotating heat exchangers that transfer heat between two air streams. They are found in thermal power plants and the HVAC systems of buildings. Energy wheels are also rotating exchangers, but transfer both heat and moisture between two air streams and are mainly found in HVAC systems. Heat/energy wheels are distinguished from other air-to-air heat/energy exchangers because of their high effectiveness values and large internal surface areas. When properly applied, heat/energy wheels are cost-effective and reduce energy consumption and environmental impacts. In new buildings, heat/energy wheels often have an immediate or very short (less than one year) payback period, while in the retrofit of existing buildings, the payback period will be a little longer (e.g., two to four years). In nearly all cases, the total life cycle costs will be lower when heat/energy wheels are applied. The reduction in life cycle costs may be in the order of 25%–50% and the energy savings may exceed the capital costs of the exchanger by an order of magnitude. An additional benefit of heat/energy wheels is that they generally have a positive impact on the environment because they reduce energy consumption and the harmful emissions associated with energy production.

REFERENCES

1. Harper, D.B.; Rohsenow, W.M. Effect of rotary regenerator performance on gas-turbine plant performance. Trans. ASME **1953**, *75*, 759–765.
2. Shah, R.K. Thermal design theory for regenerators. In *Heat Exchangers: Thermal-Hydraulic Fundamentals and Design*; Kakaç, S., Bergles, A.E., Mayinger, F., Eds.; Hemisphere: New York, 1981; 721–763.
3. Shah, R.K. Counter flow rotary regenerator thermal design procedures. In *Heat Transfer Equipment Design*; Shah, R.K., Subbarao, E.K., Mashelkar, R.A., Eds.; Hemisphere: New York, 1988; 267–296.
4. Wargocki, P.; Wyon, D.P.; Sundell, J.; Clausen, G.; Fanger, P.O. The effects of outdoor air supply rate in an office on perceived air quality, sick building syndrome (SBS) symptoms and productivity. Indoor Air **2000**, *10*, 222–236.
5. Fang, L.; Clausen, G.; Fanger, P.O. Impact of temperature and humidity on the perception of indoor air quality. Indoor Air **1998a**, *8*, 80–90.
6. Fang, L.; Wargocki, P.; Witterseh, T.; Clausen, G.; Fanger, P.O. Field study on the impact of temperature, humidity and ventilation on perceived air quality. In *Proceedings of Indoor Air'99*; Edinburgh, 1999 Vol. 2; 107–112.
7. ASHRAE Standard 90 *Energy Conservation in New Building Design*; ASHRAE: Atlanta, 1975.
8. Sundell, J. What we know, and what we don't know about sick building syndrome. ASHRAE J. **1996**, *38* (6), 51–57.
9. ANSI/ASHRAE Standard 62 *Ventilation for Acceptable Indoor Air Quality*; ASHRAE: Atlanta, 1989.
10. ANSI/ASHRAE Standard 62.1 *Ventilation for Acceptable Indoor Air Quality*; ASHRAE: Atlanta, 2004.
11. Wyon, D.P. Indoor environmental effects on productivity. In *Proceedings of IAQ 96—Paths to Better Building Environments*; Teichman, K.Y., Ed.; ASHRAE: Atlanta, 1996; 5–15.
12. Besant, R.W.; Simonson, C.J. Air-to-air energy recovery. ASHRAE J. **2000**, *42* (5), 31–42.
13. Harriman, L.G.; Plager, D.; Kosar, D. Dehumidification and cooling loads from ventilation air. ASHRAE J. **1997**, *39* (11), 37–45.
14. Harriman, L.G.; Witte, M.J.; Czachorski, M.; Kosar, D.R. Evaluating active desiccant systems for ventilating commercial buildings. ASHRAE J. **1999**, *41* (10), 28–37.
15. Besant, R.W.; Simonson, C.J. Air-to-air exchangers. ASHRAE J. **2003**, *45* (4), 42–52.
16. Pint, B.A. Stainless steels with improved oxidation resistance for recuperators. J. Eng. Gas Turbines Power **2006**, *128* (2), 370–376.
17. ASHRAE, HVAC Systems and Equipment Handbook. In *Ch 44: Air-to-Air Energy Recovery*; ASHRAE: Atlanta, 2004.
18. Simonson, C.J.; Besant, R.W. Energy wheel effectiveness. Part II—correlations. Int. J. Heat Mass Transfer **1999**, *42* (12), 2171–2185.

19. ASHRAE Standard 84 *Method of Testing Air-to-Air Heat Exchangers*; American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc.: Atlanta, 1991.
20. Simonson, C.J.; Ciepliski, D.L.; Besant, R.W. Determining the performance of energy wheels: Part I—experimental and numerical methods. ASHRAE Trans. **1999**, *105* (1), 188–205.
21. Abe, O.O.; Simonson, C.J.; Besant, R.W.; Shang, W. Effectiveness of energy wheels from transient measurements: Part II—results and verification. Int. J. Heat Mass Transfer **2006**, *49*, 63–77.
22. Abe, O.O.; Simonson, C.J.; Besant, R.W.; Shang, W. Effectiveness of energy wheels from transient measurements: Part I—prediction of effectiveness and uncertainty. Int. J. Heat Mass Transfer **2006**, *49*, 52–62.
23. Ciepliski, D.L.; Simonson, C.J.; Besant, R.W. Some recommendations for improvements to ASHRAE Standard 84–91. ASHRAE Trans. **1998**, *104* (1B), 1651–1965.
24. ANSI/ARI Standard 1060 *Standard for Rating Air-to-Air Exchangers for Energy Recovery Ventilation Equipment*; Air-Conditioning & Refrigeration Institute: Arlington, VA, 2001.
25. Besant, R.W.; Simonson, C.J. HVAC system design employing certified air-to-air energy recovery equipment. In *Proceedings (CD) of the Worldwide CIBSE/ASHRAE Conference, Building Sustainability, Value and Profit*, *eight pages*, *Edinburgh*, *Scotland*; September 24–26, 2003.
26. Phillips, E.G.; Fisher, D.; Chant, R.; Bradley, B. Freeze-control strategy and air-to-air energy recovery performance. ASHRAE J. **1992**, *34* (12), 44–49.
27. Simonson, C.J.; Besant, R.W. Heat and moisture transfer in energy wheels during absorption, condensation and frosting conditions. ASME J. Heat Transfer **1998**, *120* (3), 699–708.
28. Asiedu, Y.; Besant, R.W.; Simonson, C.J. Wheel selection for heat and energy recovery in simple HVAC ventilation design problems. ASHRAE Trans. **2004**, *110* (1), 381–398.
29. Asiedu, Y.; Besant, R.W.; Simonson, C.J. Cost effective design of dual heat and energy recovery exchangers for 100% ventilation air in HVAC cabinet units. ASHRAE Trans. **2005**, *111* (1), 857–872.
30. Häkkinen, T.; Tattari, K.; Klobut, K.; Nyman, N.; Saari, M.; Laine, J. Environmental Impacts of Building Services Devices and Systems, Basics of Assessment, and Environmental Declarations; *The Finnish Development Centre for Building Services*; Report 11; Helsinki; Finland (in Finnish) 1999.
31. Nyman, M.; Simonson, C.J. Life cycle assessment of residential ventilation units in a cold climate. Build. Environ. **2005**, *40* (1), 15–27.
32. Nyman, M.; Simonson, C.J. Life cycle assessment (LCA) of air-handling units with and without air-to-air energy exchangers. ASHRAE Trans. **2004**, *110* (1), 399–409.

Heat Exchangers and Heat Pipes

Greg F. Naterer
University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Abstract
This entry provides an overview of the design and opertion of heat exchangers and heat pipes. Different types of heat exchangers are discussed, including concentric-tube, cross-flow, and shell-and-tube heat exchangers. Methods of analysis are briefly discussed, with correction and geometrical factors for complex configurations. Condensers and evaporators are described in more detail. Heat pipes are additional types of heat exchange devices that use a porous wick material in a tube structure. The performance and operation of heat pipes are discussed. Limitations of wicking, entrainment, sonic, and boiling conditions are outlined, with respect to their effects on heat pipe performance.

INTRODUCTION

Heat exchangers are devices allowing energy exchange between fluid streams at different temperatures. They are commonly used in many applications, including power generation, energy storage, air conditioning systems, materials processing, and various others. Some common configurations include concentric-tube, cross-flow, and shell-and-tube heat exchangers (see Figs. 1 and 2). Another widely used device for heat exchange is a heat pipe, for applications such as spacecraft thermal control and thermal energy management in microelectronic assemblies. This entry explains the fundamental operating principles of heat exchangers and heat pipes.

In a concentric-tube heat exchanger, an internal fluid flows through the inner tube, and an external flow passes through the annular region between the inner and outer tubes. In a parallel-flow configuration, the outer fluid flows in the same direction as the inner flow. Otherwise, if the outer fluid flows in a direction opposite to the inner flow, then it is called a counter-flow heat exchanger. Analysis of heat exchangers usually requires empirical or advanced computer simulation techniques.

A cross-flow heat exchanger consists of an outer flow passing across tubes carrying fluid that flows in a direction perpendicular to the cross-flow. An example of this type of heat exchanger is found in a car radiator, where the outer and inner flows consist of air and water, respectively. The tubes are often covered with fins or other annular attachments to enhance the rate of heat transfer between the different fluid streams. For systems having fluid streams separated from one another by fins or baffles, the configuration is called unmixed, whereas a mixed configuration permits complete mixing of the fluid streams in the external cross-flow.

Another common type of heat exchanger is a shell-and-tube heat exchanger. In this case, the configuration consists of an outer shell pipe where fluid enters through one end, passes across internal tubes carrying a fluid at a different temperature, and exits through the other end. Baffles are often placed perpendicular to the inner tubes to enhance mixing and turbulence of the outer fluid stream. Baffles refer to perforated plates that obstruct some region of the outer flow while directing the flow around the remaining uncovered sections. Condensers in power plants are common examples of shell-and-tube heat exchangers. In these condensers, the outer flow is steam that condenses and leaves as water following heat exchange with the inner tubes carrying cold water.

Heat transfer occurs between fluids at different temperatures or phases. In the case of heat exchange between fluids of the same phase, but at different temperatures, the heat gained by the colder fluid stream balances heat lost by the hotter fluid stream, minus any external heat losses to the surroundings (often assumed negligible). Spacing and packing of tubes within heat exchangers varies with different applications. The surface area density of surfaces characterizes this packing, based on the number and diameter of tubes within the heat exchanger. The packing is often expressed in terms of a hydraulic diameter of tubes, D_h. A typical range is $0.8\ \text{cm} < D_h < 5\ \text{cm}$ for shell-and-tube heat exchangers, while $0.2\ \text{cm} < D_h < 0.5\ \text{cm}$ for automobile radiators and $0.05\ \text{cm} < D_h < 0.1\ \text{cm}$ for gas turbine regenerators. For heat exchange in biological systems (such as human lungs), the range is typically $0.01\ \text{cm} < D_h < 0.02\ \text{cm}$.

HEAT EXCHANGER ANALYSIS

Heat exchange between two fluid streams in a concentric-tube heat exchanger depends on the overall heat transfer

Keywords: Tubular heat exchanger; Cross-flow heat exchanger; Shell-and-tube heat exchanger; Condensers and evaporators; Heat pipe; Wicking limit; Entrainment limit; Sonic limit.

Heat–Inde

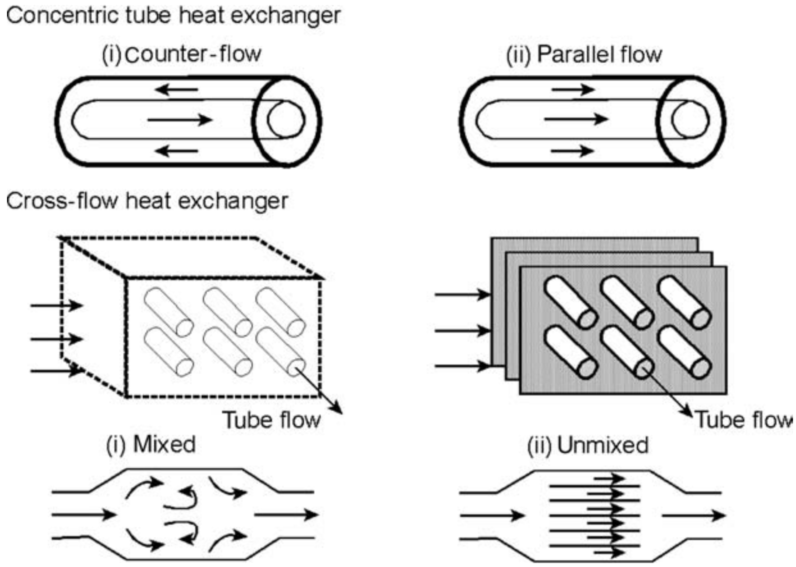


Fig. 1 Types of heat exchangers.

coefficient between both streams, including thermal resistances due to convection, fouling (due to fluid impurities such as rust formation), and conduction through the pipe wall. Frequent cleaning of the heat exchanger surfaces is needed to reduce and minimize the adverse effects of fouling, such as an increased pressure drop and reduced heat transfer effectiveness. Finned surfaces produce additional thermal resistances in a heat exchanger. These effects are often modeled based on the surface efficiency of a finned surface, which includes the efficiency for heat exchange through the fin, as well as heat transfer through the base surface (between the fins).

Using analytical methods for basic parallel-flow configurations of heat exchangers, the total heat transfer from a hot fluid stream to a cold stream between an inlet (1) and outlet (2) can be readily determined based on given (or measured) temperature differences, ΔT_1 and ΔT_2. A similar analysis can be derived for both parallel and counter-flow heat exchangers. In a parallel-flow heat exchanger, the highest temperature difference is encountered between the two incoming fluid streams. In the streamwise direction, heat transfer from the hot stream to the cold stream reduces the temperature difference between the fluids. But in a counter-flow heat exchanger, the temperature difference increases in the flow direction, since the temperature of the incoming cold fluid stream increases due to heat transfer from the hot stream flowing in the opposite direction. A counter-flow heat exchanger is generally considered to be more effective, since a smaller surface area is required to achieve the same heat transfer rate (assuming equivalent heat transfer coefficients between the fluid streams).

Predicted results for parallel-flow and counter-flow heat exchangers can be extended to more complicated geometrical configurations, such as cross-flow and shell-and-tube heat exchangers, after multiplying the heat transfer rate by a correction factor, F, which is usually based on experimental data to account for baffles and other geometrical parameters. The value of F depends on the type of heat exchanger. For example, $F=1$ for a one-shell-pass, one-tube-pass heat exchanger. For more complex arrangements, values of F are usually tabulated

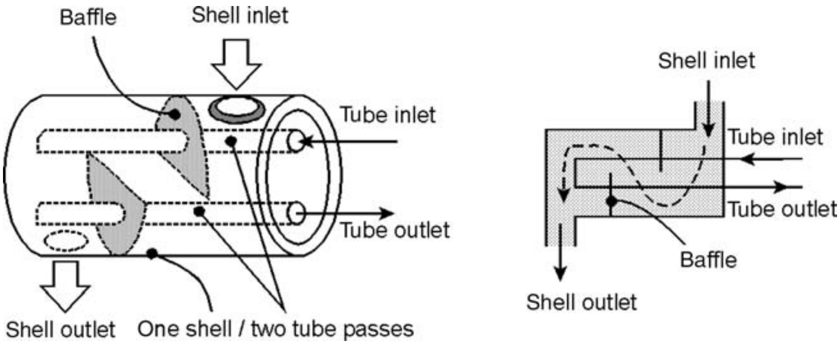


Fig. 2 Shell-and-tube heat exchanger.

and graphically depicted. For example, results of correction factors for a variety of heat exchanger configurations have been presented and graphically illustrated by Bowman et al.[1] Incropera and Dewitt,[5] and others. The *Standards of the Tubular Exchange Manufacturers Association* (6th edition, New York, 1978) provides additional results in terms of algebraic expressions or graphical representations.

The competing influences of pressure drop and heat exchange are important considerations in heat exchanger design. Larger rates of heat transfer can usually be achieved by increasing the packing of tubes within the heat exchanger or using baffles or other heat enhancement devices. But this comes at the expense of an increased pressure drop, which is disadvantageous due to additional pumping power required to move the fluid through the system at a specified mass flow rate. On the other hand, fewer heat exchange tubes can lead to a smaller pressure drop, but often at the expense of lower heat transfer, in comparison to a design with a high surface area density. Thus, optimization serves to provide an effective balance between the heat exchange and pressure losses. Empirical correlations are often used to predict the pressure drop in heat exchangers, in terms of additional parameters such as fluid velocity, the ratio of the free-flow area of the finned passages to the frontal area of the heat exchanger, friction factor, flow length, and flow passage hydraulic radius (total heat exchanger volume divided by the total heat transfer surface area). Values of friction factors have been documented extensively by Kays and London[6] for a variety of heat exchangers, including finned and various tubular configurations.

Plate-to-fin heat exchangers are commonly used in applications involving heat exchange between gas-to-gas streams. For example, air-to-air heat exchangers are typically based on plate-to-fin arrangements. These fins are classified into various types, such as plain fins, strip fins, pin fins, perforated fins, and others. Detailed design data involving these types of heat exchangers are outlined in Kays and London.[6]

CONDENSERS AND EVAPORATORS

Other common types of heat exchangers are condensers and evaporators, which are two-phase heat exchangers used in various engineering systems, such as power-generation and refrigeration systems. When a fluid stream evaporates or condenses within a heat exchanger and experiences a change of phase, then it is usually more useful to evaluate enthalpy (rather than temperature) in energy balances for heat exchanger analysis. The fluid temperature can remain nearly constant during the phase change, even though heat is transferred between the fluid streams. When calculating an enthalpy difference, the heat transfer analysis would include both the latent and sensible heat portions of the energy transfer between different fluid streams in the heat exchanger.

Unlike previously discussed heat exchangers involving single phase flows, a main difficulty in the analysis of condensers and evaporators is the range of phase-change regimes experienced by the fluid stream. The heat transfer coefficient depends on the local phase fraction, which varies throughout the flow path, so the heat transfer coefficient becomes position dependent. Unfortunately, the phase distribution is generally unknown until the flow field solution is obtained. Thus, a systematic procedure is needed to analyze heat transfer processes in condensers and evaporators.

The energy balances involve enthalpy to accommodate the latent heat of phase change, as well as sensible heat portions of the heat exchange. The fluid possibly undergoes various phase-change regimes, so the tube length is usually subdivided into discrete elements, and energy balances are applied individually over each element. After the enthalpy in a particular element is computed, its value may exceed the saturated vapor value of enthalpy at the flow pressure. In this case, the fluid enthalpy can be used to calculate the temperature of the superheated vapor based on thermodynamic property tables or computer-generated tabulation of the superheated property values. Alternatively, the specific heat of the fluid can be used to predict the temperature change corresponding to the enthalpy difference. This approach assumes a locally constant value of the specific heat, which requires a sufficiently small temperature change between elements. This method can be used when the fluid exists entirely as a superheated vapor, as the vapor specific heat at the mean temperature can be used.

However, if the enthalpy is computed to be less than the saturated vapor enthalpy, then the quality (mass fraction of vapor in the element) is needed. An updated estimate of the convection coefficient can be calculated based on the phase fraction and corresponding flow regime. Additional empirical factors (such as a transition factor) are often used to identify the flow regime and appropriate correlation for heat transfer. Then an updated overall heat transfer coefficient can be determined from the phase fraction and flow regimes.

A typical numerical procedure for analyzing two-phase heat exchangers can be summarized as follows: a boundary condition is first applied within the initial element of the tube. Then, a suitable forced convection correlation is used up to the element in which phase change is first realized. An appropriate heat transfer correlation can then be selected for that phase change regime. Near the points of saturated vapor or saturated liquid, a suitable single-phase correlation can be used with property values along the saturated liquid and vapor lines. Then, this procedure can be repeated for each element in the domain. For either condensing or boiling problems, a similar procedure can be adopted. In the latter case (boiling), a two-phase flow map would typically be utilized to identify the flow regime based on the computed phase fraction. This mapping would distinguish between flow regimes, such as the wavy, annular, and slug flow regimes.

Various design features and aspects of maintenance are important in terms of the effective performance of condensers and evaporators. Tubes should be readily cleanable on a regular basis, either through removable water heads or other means. Larger flow rates within the heat exchanger can reduce fouling (buildup of scale and dirt on the walls), reduce service, and extend the life of the heat exchanger. Furthermore, higher operating efficiencies can be achieved by placement of tubes in stacks with metal-to-metal contact between fins to permit better drainage of the condensate. Less thermal resistance occurs when less liquid accumulates on the fins, thereby improving the thermal efficiency of heat exchange. Also, a light and compact design is beneficial, as it requires less space and reduces difficulties in installation and moving, while often reducing costs associated with maintenance.

Another important factor in proper operation of condensers and evaporators is safety. For systems operating at pressures different from the surrounding ambient (atmospheric) pressure, leakage can occur. In maintenance procedures, the specific points of leakage can be detected and repaired by a basic method of soap or detergent brushed onto the surfaces where leakage is suspected, thereby generating bubbles to indicate leakage points. Alternatively, pressurizing the system and recording changes in pressure over time can indicate the tightness of a system (but not necessarily the location of leakage). Certain chemical leaks can be detected individually. For instance, sulfur dioxide can be detected by the white smoke forming when ammonia is brought into close contact with the leakage point.

Operating materials must be properly selected in conjunction with the working fluids. Most refrigerants can be used well under normal conditions with most metals (such as steel, aluminum, and iron), but some materials and liquids should never be used together. An example is methyl chloride fluid with aluminum shells and tubes, which can produce flammable gas byproducts. Also, the tensile strength, hardness, and other properties of exposed materials must be fully considered under all operating conditions. Effects of certain plastic materials on refrigerant liquids can often be difficult to predict, particularly due to the rapid rise in the number and types of polymer materials. An effective overall design of condensers and evaporators requires a thorough investigation of both thermal and materials engineering aspects.

HEAT PIPES

A heat pipe is a closed device containing a liquid that transfers heat under isothermal conditions. It operates through vaporization of liquid in an evaporator, transport and condensation of the vapor, and return flow of liquid by capillary action through a wick structure back to an evaporator. Due to geometrical requirements, the adiabatic section is designed to fit within spacing limitations of the heat pipe. Adiabatic implies zero heat transfer, as in a well-insulated section. Thermal energy from the external source is transferred to the working fluid in the heat pipe at the evaporator section. At the end of the heat pipe, a buffer volume may be constructed to enclose a non-condensable gas (such as helium or argon) for controlling the operating temperature, based on control of pressure within the inert gas. Flow of vapor occurs through the core interior region of the heat pipe at high velocities to the condensing section (up to 500 MPH in some cases).

Along the inner wall of the container of the heat pipe, a porous wick material with small, random interconnected channels is constructed for capillary pumping. The pores in the wick act as a capillary "pump," which acts analogously to regular pumping action on fluids in pipes by pumps. The wick provides an effective means of transporting liquid back to the evaporator through surface tension forces acting within the wick. Also, it serves as an effective separator between vapor and liquid phases, thereby allowing more heat to be carried over greater distances than other pipe arrangements (Fig. 3).

Various applications utilize heat pipes, including heating, ventilating, and air conditioning (HVAC) heat recovery systems; microelectronics cooling; and spacecraft thermal control. Heat pipes in air-to-air HVAC heat recovery systems allow effective storage of thermal energy contained in exiting combustion gases. Heat pipes offer key advantages over conventional techniques, including low maintenance (no moving parts), long life, and cost savings. Another example involves microelectronics cooling. Heat pipes can be up to 1000 times more conductive than copper (at the same weight). Examples include laptop computers, as well as telecommunications equipment, which have adopted heat pipes with success in their thermal designs. Also, heat pipes appear in several spacecraft thermal control applications. Heat pipes have been used in satellites to transfer heat generated by electronic equipment to radiation panels that dissipate heat into space. Another application is tubing in satellites, which provides effective control of temperatures required for reliable performance of electrical components on the satellite.

In the evaporator section of a heat pipe, heat is transferred by conduction from the energy source through the container wall and wick-to-liquid matrix to the vapor-to-liquid interface. Then, liquid is evaporated at the vapor-to-liquid interface, and heat transfer occurs by convection of vapor (laminar or turbulent) from the evaporator to the condenser. The temperature of the vapor is approximately the average between the source and sink temperatures at the ends of the heat pipe. Following condensation of vapor

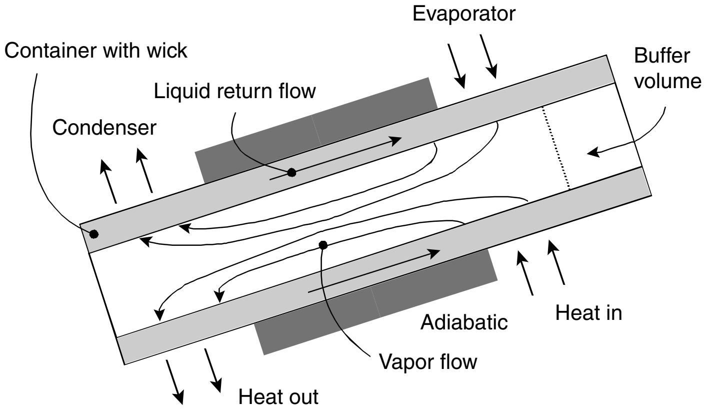


Fig. 3 Schematic of a heat pipe.

at the condenser end, heat transfer by conduction occurs through the wick-to-liquid matrix and container wall to the heat sink. Finally, liquid condensate returns to the evaporator through the wick structure (generally laminar) return flow.

The working fluid and wick type are important design factors in a heat pipe. The working fluid should have a high latent heat of evaporation, high thermal conductivity, high surface tension, low dynamic viscosity, and suitable saturation temperature. Also, it should effectively wet the wick material. Some typical examples of working fluids are water or ammonia at moderate temperatures, or liquid metals such as sodium, lithium, or potassium at high temperatures (above 600°C). In a typical example of a heat pipe with water as the working fluid and a vessel material of copper–nickel, an axial heat flux of about 0.67 kW/cm^2 at 473 K and a surface heat flux of about 146 W/cm^2 at 473 K are generated. For moderate operating temperatures (i.e., 400–700 K), water is the most suitable and widely used working fluid in heat pipes.

It is interesting to compare heat pipes with solid materials in terms of an effective thermal conductivity. Consider the axial heat flux in a heat pipe, using water as the working fluid at 200°C, in comparison with the heat flux in a copper bar (10 cm long) experiencing a temperature difference of 80°C. For the copper bar at 200°C, the heat flux per unit area can be estimated at about 0.03 kW/cm^2. In contrast, the axial heat flux for a water heat pipe under the given conditions is about 0.67 kW/cm^2, so the heat pipe transfers heat at a rate more than 20 times higher than a copper rod. This example indicates that heat pipes can be beneficial by providing a higher effective thermal conductivity of material.

In a heat pipe, proper liquid circulation is maintained within the heat pipe as long as the driving pressure (capillary forces) within the wick exceeds the sum of frictional pressure drops (liquid and vapor) and the potential (gravity) head of the liquid in the wick structure. Capillary action in the wick arises from surface tension forces at the liquid interface. In the wick structure, a liquid flow is generated by the capillary action due to liquid entrainment within the wick structure. Spatial differences of capillary pressure (due to different curvatures of liquid menisci) induce the capillary flow. Another pressure drop occurs when liquid flows (generally laminar) through grooves in the wick from the condenser back to the evaporator. This liquid flow can be analyzed through from Darcy's law for laminar flow in a porous medium. The wick structures can have a wrapped screen along the inner wall of the heat pipe or screen-covered grooves, in order to improve their operating performance. Examples of common wick materials are copper foam, copper powder, felt metal, nickel fiber, and nickel powder.

A vapor pressure drop arises since vapor drag in the core region may impede liquid flow in the grooves of the wick at high vapor velocities. An expression for this pressure drop can be determined from standard methods of fluid mechanics. This pressure drop is usually a small contribution to the overall force balance, since the vapor density is much smaller than the liquid density. Also, effects of vapor drag can be reduced by covering the grooves in the wick structure with a screen.

The equilibrium design condition of a heat pipe is determined from a balance of the previous pressure forces, including a gravity term, which can be positive (gravity assisted) or negative. A positive gravity head implies that the evaporator is above the condenser. The design condition is significant because operating beyond the maximum capillary pressure can dry out the wick and produce a "burnout" condition.

Wicking, entrainment, sonic, and boiling limitations have significant effects on the performance of heat pipes. A wicking limitation occurs on the axial heat flux, due to a maximum flow rate through the wick for a maximum capillary pressure rise. Entrainment refers to droplets entrained by the vapor from the liquid return flow. Under certain operating conditions, the vapor velocity may become sufficiently high to produce shear force effects on the liquid return flow from the condenser to the evaporator. It is possible that waves can be generated on the liquid surface and droplets may be entrained by the vapor flow, since there would be inadequate restraining forces of liquid surface tension in the wick.

Another factor is the sonic limitation. During conditions of startup from near-ambient conditions, a low vapor pressure within the heat pipe can lead to a high resulting vapor velocity. If the vapor velocity approaches sonic speed, a choked condition within the pipe limits the axial heat flux. This sonic limit and other previous depend on the fluid operating temperature. Heat flux limits generally increase with evaporator exit temperature due to the effect of temperature on the speed of sound in the vapor. For example, the heat flux limit for sodium increases from 0.6 kW/cm^2 at 500°C to 94.2 kW/cm^2 at 900°C. For liquid potassium, the heat flux limit is 0.5 kW/cm^2 at 400°C (evaporator exit temperature), and the limit increases to 36.6 kW/cm^2 at 700°C. In high-temperature applications, lithium can be used. Its heat flux limit ranges between 1.0 kW/cm^2 at 800°C and 143.8 kW/cm^2 at 1300°C.

In contrast to the limitations on the axial heat flux, the boiling limitation involves the radial heat flux through the container wall and wick. The onset of boiling within the wick interferes with and obstructs the liquid return flow from the condenser. Boiling within the wick may cause a burnout condition by drying out the evaporator containment. Recent advances in heat pipe technology are developing innovative techniques for dealing with this thermal limitation and enhancing the overall capabilities of heat pipes. Additional references in the topic of heat pipe analysis are given by Kreith and Bohn,[7] Hewitt et al.[4] Dunn and Reay,[3] and Chi.[2]

REFERENCES

1. Bowman, R.A.; Mueller, A.C.; Nagle, W.M. Mean temperature difference in design. Trans. ASME **1940**, 62.
2. Chi, S.W. *Heat Pipe Theory and Practice*; Hemisphere: Washington, DC, 1976.
3. Dunn, P.D.; Reay, D.A. *Heat Pipes*, 3rd Ed.; Pergamon: New York, 1982.
4. Hewitt, G.F., Shires, G.L., Polezhaev, Y.V., Eds., *International Encyclopedia of Heat and Mass Transfer*, CRC Press: Boca Raton, FL, 1997.
5. Incropera, F.P.; DeWitt, D.P. *Fundamentals of Heat and Mass Transfer*, 3rd Ed.; Wiley: New York, 1990.
6. Kays, W.M.; London, A.L. *Compact Heat Exchangers*, 3rd Ed.; McGraw-Hill: New York, 1984.
7. Kreith, F.; Bohn, M.S. *Principles of Heat Transfer*, 6th Ed.; Brooks/Cole Thomson Learning: Pacific Grove, CA, 2001.

Heat Pipe Application

Somchai Jiajitsawat

John Duffy

John W. Wang

Department of Mechanical Engineering, University of Massachussetts—Lowell, Lowell, Massachussetts, U.S.A.

Abstract

Heat pipes have been known to be effective heat transfer devices for well over 70 years. They have been widely employed in various applications throughout industry. This article addresses various heat pipe applications in many disparate areas, such as aerospace, medicine, agriculture, transportation, and the automotive industry.

Examples in the application of heat pipe technology are given in heat exchangers, in electronic components, in medical devices, in ovens and furnaces, and in solar thermal applications.

INTRODUCTION

Heat pipes are heat transfer devices that have been widely used in numerous applications for at least 70 years. The concept of a "two-phase heat transfer device capable of transferring large quantities of heat with a minimal temperature drop" was first introduced by R. S. Gaugher in 1942. Since then, heat pipes have been employed in various industries, especially in the aerospace industry, in the various engineering disciplines, in medicine, in transportation, and in agriculture. The use of heat pipe technology is ubiquitous in heat exchangers, in the cooling of electronic components, and in solar thermal applications.

Heat pipes are essentially heat transfer devices, transferring heat from one area to another. In principle, heat pipes operate in a closed container, constituting a sealed environment lined with a wicking material saturated with working fluid. The pipes operate in a two-phase cycle in which the working fluid is either in a liquid state or exists in vapor form. Typically, in situations in which a cooling effect is desired, heat is absorbed in some manner at the evaporator end of the heat pipe and rejected or dissipated at the condenser end of the device as the vapor condenses to liquid. Contrariwise, when a warming or heating result is desired, heat is applied to the liquid at the condenser end of the heat pipe and released from the evaporator end, thus delivering heat or warmth at that point.

Heat pipe applications can be classified in four main categories: separation of heat source and sink, temperature equalization, temperature control, and thermal diode. Due to their high thermal conductivity, heat pipes can efficiently transport heat from source to sink without regard to heat sink space requirements. Another advantage of the high thermal conductivity is the ability to provide an effective method of temperature equalization. For example, a heat pipe located between two opposing faces of an orbiting platform can maintain constant temperature in both faces of the platform, thus minimizing thermally induced structural stresses. The resultant temperature control is the product of the capability of heat pipes to transport large quantities of heat without significant increases in the operating temperature of the heat pipe. A thermal diode is a heat pipe function for situations in which heat has to be transferred efficiently in one direction and heat flow in the reverse direction has to be prevented.

This entry will focus on the various areas of heat pipe applications. A brief account of various applications of heat pipe technology in each area will follow.

AEROSPACE

Heat pipes enjoy wide application in the area of spacecraft cooling and temperature stabilization. They enjoy the distinct advantages of low weight, essentially zero maintenance, and superior reliability over other devices. Structural isothermal operation is an important goal in aerospace operations. Thermal stress occurring from heat inequalities is a critical issue in many orbiting astronomical experiments. While in orbit, for example, an observatory may be fixed on a single point such as a star. Consequently, one side of the capsule will be subjected to intense solar radiation while the other is not. Heat pipes in this situation are used to transport heat from the side facing the sun to the cold side away from it, thus

Keywords: Heat pipe(s); Heat transfer; Heat sink; Heat pipe application; Heat-pipe heat exchanger; Heat pipe cooling; Heat pipe furnace; Heat pipe oven.

equalizing the temperature of the structure.[1] Because of the demand for reducing spacecraft costs while maintaining high-performance characteristics of the spacecraft bus, the mass of satellites has to be minimized. Thus, the reduction of space cooling by increasing heat density is an important challenge. To achieve this, heat pipes are commonly used to affect heat transfer and heat redistribution functions in the microsatellites.[2]

HEAT EXCHANGERS

Because of flexibility in design, heat pipes can easily be utilized as heat exchangers inside sorption and vapor-compression heat pumps, refrigerators, and other types of heat transfer devices. For example, heat pipe heat exchangers (HPHE) used for controlling humidity in air conditioning systems have been exhaustively researched.[3] The purpose of the HPHE is to recover heat from warm outdoor air to reheat the dewpoint airstream, thereby minimizing heating costs. Also, the evaporator of the HPHE acts as a precooler to the warm outdoor air before it reaches the air conditioner, thus enhancing the effectiveness of the cooling coil. When the condenser of the HPHE is used as a reheater to heat the airstream coming from the cooling coil, the relative humidity can more easily be reduced to the comfort zone, below 70%.

This HPHE concept can also be applied in various industries, such as metallurgy, power plants, and oil refining. For instance, in the thermal power industry, heat pipe air preheater/exchangers (see Fig. 1) are used to preheat the air used in the combustion of fuel in the boiler. This is achieved by retrieving the energy available in the exhaust gases. In this application, two ducts are held in parallel, with a common wall. Hot flue gases are made to flow through one duct while cold combustion air is directed through the other duct, causing the cold air to be "preheated."

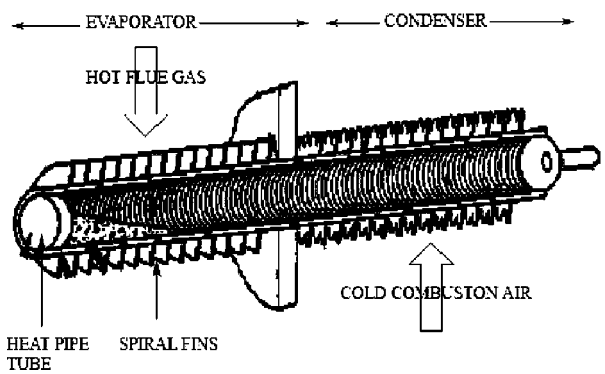


Fig. 1 Heat pipe air preheater.
Source: From Elsevier (see Ref. 4).

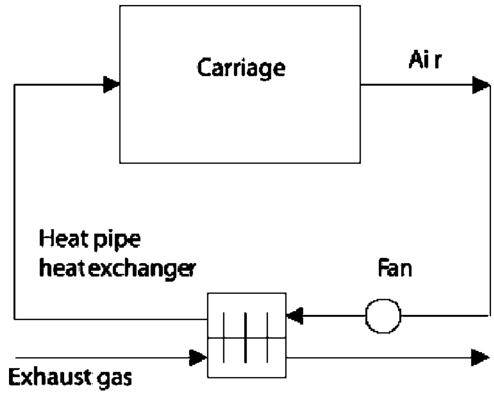


Fig. 2 Diagram of heating system.
Source: From Elsevier (see Ref. 5).

Heat pipe exchangers are also used to recover the thermal energy otherwise lost in exhaust gases of ground transportation vehicles—in buses, in vans and trucks, and in passenger cars. The heat thus recovered can be used for heating the cabins or the passenger areas in cold weather. The schematic diagram of the heating system used is in Fig. 2.

ELECTRONIC COMPONENTS

As of the present, one of the largest applications of heat pipe technology is the cooling of electronic components such as central processing units (CPUs), circuit boards, and transistors. For example, the CPU, one of the most important parts of any PC, is becoming increasingly more compact, faster, and more efficient. This leads to higher heat density, resulting in increasing CPU temperature, threatening a shortened life of the chip, or resulting in malfunction or failure. A conventional method used to keep a CPU from overheating is to use an extruded aluminum heat sink. This is an efficient method in terms of unit price, weight, and performance. However, as computers become smaller, the ability of

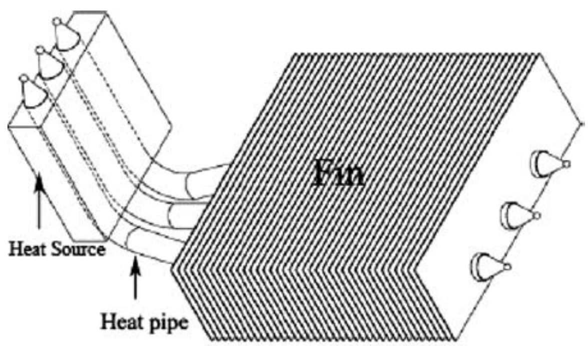


Fig. 3 Heat pipe cooling where heat source locates in the top and fin in the bottom.
Source: From Elsevier (see Ref. 6).

the heat sink to effect cooling becomes insufficient to meet demands, chiefly because of insufficient space. The use of heat pipes in such situations has marked advantages (see Fig. 3).

OVENS AND FURNACES

One of the earliest applications of heat pipe technology was in conventional baking ovens. Previously, flames were applied to the firebrick lining the oven which in turn transmitted the heat to the items baked in the oven. Unfortunately, the baked items were contaminated by smoke and soot and other combustion products produced by the flames.

In improved models using heat pipes, the transfer of heat to the oven was effectuated by the evaporation and condensation of the working fluid in the closed heat pipe system, which transferred the heat, but not the combustion products. The use of heat pipes conferred two other advantages: a more uniform oven temperature, and a savings of up to 25% of the fuel normally consumed.

Recently, in the fine-chemical industry, higher temperature requirements have been mandated for the spray-drying of powdered materials, which require a hot air temperature of 450°C–600°C or higher. It is difficult to heat the air to such high temperatures with conventional heat exchange equipment. Also, if the flue gas of the fuel were to be applied directly, pollutants might be carried with it, rendering the product quality unacceptable. Fig. 4 shows a newly developed heat pipe high-temperature hot air furnace that overcomes these objections.

SOLAR THERMAL APPLICATIONS

Due to a shortage in the supply of fossil fuel, renewable energy sources have become one of many alternative solutions. Solar energy is a renewable energy resource, and it is abundant. It has been used in a variety of applications in generating electricity, in cooling, and even in heating. The following solar applications employing heat pipe technology are discussed.

Solar Distillation

Recently, research has been directed to the discovery of alternative sources of energy. Ethanol is one of the key renewable fuels utilized to partially replace petroleum products. It is produced from agricultural products such as cassava, corn, sugarcane, molasses, and cereal grains. Solar distillation is one such technique. It is used to generate ethanol and results in the consumption of only a small amount of fossil fuel compared with conventional processes. Evacuated heat pipe collectors (see Fig. 5) are used in this system to generate the heat required for the distillation process.

Solar Cooking

Preparation of foods requires some amount of energy. Wood fuel and agricultural residues are the main energy sources for cooking in developing countries, normally accounting for 50%–90% of all energy consumed in those countries. However, the rate of wood fuel consumption exceeds its replacement, and this results in deforestation, pollution, soil erosion, global warming, and a worsening of the welfare of millions of people all over the world. Although electric cooking is comparatively convenient, its production from fossil fuel has cumulative consequences resulting in the emission of high quantities of carbon dioxide and sulfur dioxide into the atmosphere. The use of solar cooker technology is an alternative solution to the fuel problem. Heat pipes are utilized for solar cookers in transporting heat from the heat source to its destination, as shown in Fig. 6.

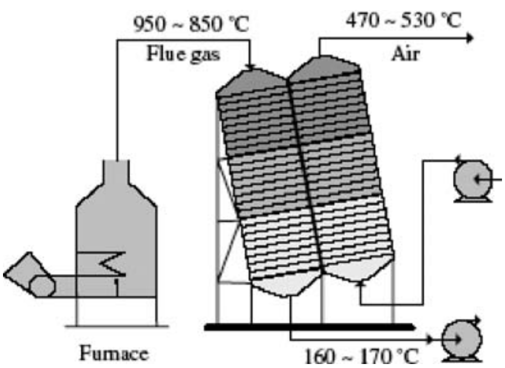


Fig. 4 High-temperature heat pipe hot air furnace.
Source: From Elsevier (see Ref. 7).

Fig. 5 Evacuated heat pipe collectors.
Source: From Elsevier (see Ref. 8).

Heat–Inde

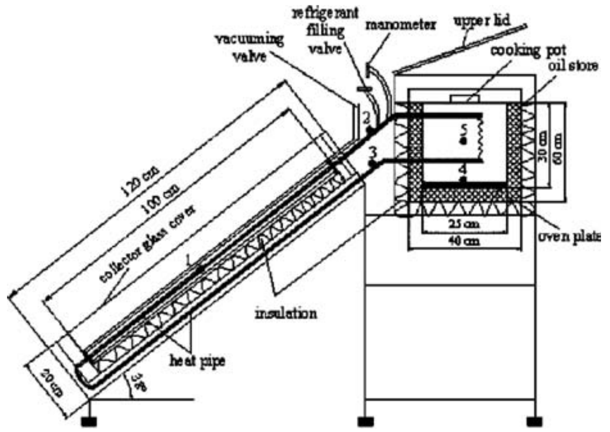


Fig. 6 Solar cooker integrated with heat pipes.
Source: From Elsevier (see Ref. 9).

AGRICULTURE

In the process of agricultural production, a soil-free planting system is widely accepted as an efficient means of improving production. There are two types of soil-free planting systems: hydroponic and aeroponic systems. The aeroponic system is a soil-free system in which the roots of the plants are completely or partially exposed to air in a plant chamber. The plants are anchored in holes atop a panel of polystyrene foam. Inside the chamber, a fine mist of nutrient solution is sprayed onto the roots.

Evaporative cooling and refrigerant systems are frequently employed to control greenhouse and nutrient temperatures. For conventional aeroponic system, an exhaust fan and an evaporative cooling unit are used to control the greenhouse temperature. This is considered to be uneconomical due to their need for high electrical energy consumption, which accounts for 33% (fan) and 25% (cooling unit) of the total energy consumed. In addition, when the nutrient solution passes through the chamber, its temperature will be unavoidably increased due to the process of plant metabolism. Normally, a refrigeration system is needed to maintain the temperature of the nutrient solution at a specific level. This results in increased energy consumption. Thus, a device is needed to reduce the energy consumption that occurs during evaporative cooling of the greenhouse and the refrigeration process. The heat pipe is a good solution to the problem. Its application in an aeroponic system is shown in Fig. 7. The evaporator section of the heat pipe removes heat from the nutrient solution that flows out from the chamber. Therefore, the outlet temperature of the solution from the evaporator section is decreased, because heat is absorbed by the heat pipe. Similarly, the heat pipe can be used to reduce the heat generated in the greenhouse, thus reducing the cooling load of evaporative cooling.

MEDICAL APPLICATIONS

One of the more practical heat pipe applications in medicine is in the heat pipe cooler (see Fig. 8). This cooler has a copper–methanol heat pipe with a copper sintered powder wick and four Peltier elements, cooled by a water heat exchanger. It is used for noninvasive treatment of an inflamed vagina, rectum, or pelvis and for some common gynaecological and rectal conditions. The main component of the cooler used in this application is the heat pipe (1). The condenser of the heat pipe is thermally contacted with the cold surfaces of Peltier elements (2). The evaporator of the heat pipe is cooled by the water in the heat exchanger (3). Peltier elements and this heat exchanger are located inside the handle of the cooler (4). During the medical treatment, the heat pipe cooler is introduced into the rectum or vagina, cooling the mucous membrane. The cooling effect is not only positive for the above-mentioned organs, but also effective in the indirect cooling of reflex ganglions in conditions such as chronic gastritis, ulcerative colitis, acute gastritis, and gastric and duodenal ulcers.

Electrosurgery is commonly used to cauterize, cut, or coagulate tissue. In typical electrosurgical devices, Radio Frequency electrical energy is applied to the tissue being treated. Local heating of the tissue occurs. By varying the power output and the type of electrical waveform, it is

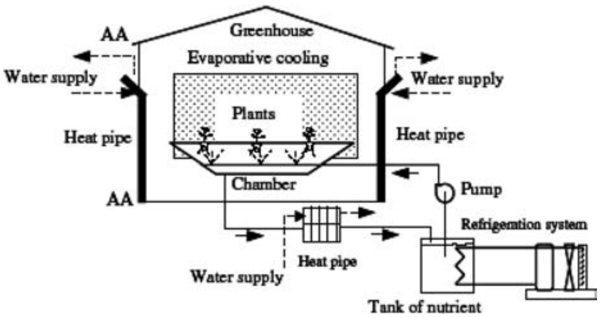


Fig. 7 Aeroponic system with heat pipe.
Source: From Elsevier (see Ref. 10).

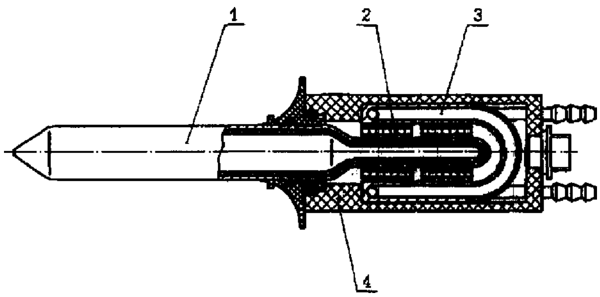


Fig. 8 Heat pipe cooler.
Source: From Elsevier (see Ref. 11).

possible to control the extent of heating and, thus, the resulting surgical effect. There are two types of electrosurgery: bipolar and monopolar. For bipolar, the device has two electrodes. The tissue being treated is placed between the electrodes, and the electrical energy is applied across the electrodes. In the monopolar device, the electrical energy is applied to a single electrode applied to the tissue, with a grounding pad placed in contact with the patient. The energy passes from the single monopolar electrode through the tissue to the grounding pad.

During this surgery, the tissue being treated tends to stick to the electrode, resulting in various complications. First, the tissue stuck to the electrode has a high resistance that hinders delivery of power to the tissue. While performing a procedure, a surgeon must periodically remove the device from the patient and clean it before continuing. Due to tissue's adhering to the electrode, when the electrode is removed, the portion of the tissue adhering to the electrode is torn away, resulting in bleeding. Therefore, cooling of the surgical site during electrosurgery is highly desirable. To achieve this, a heat pipe is introduced to maintain a constant temperature of 80°C at the electrode-tissue interface when the electrosurgical device is used.[12]

TRANSPORTATION SYSTEMS

Several heat pipe applications are used to improve the safety and reliability of air, surface, and rail transportation systems. In Siberia, during the long winter, snow may fall at the rate of 1000–1500 mm per day with the ambient air at −5 to −7°C. Heat pipe technology is used to heat the rails to prevent ice formation, as shown in Fig. 9. The condenser end of the heat pipe is located at the point of contact between the heat pipe and the rail.

In Virginia, a heated bridge[13] exists on Route 60 over the Buffalo River in Amherst County in the eastern foothills of the Blue Ridge Mountains. Road conditions during winter storms there can often be treacherous. An anti-icing heating system was designed and fabricated to obviate the situation. The system consists of approximately 2 mi of steel piping, including 241 heat pipes embedded in the concrete deck and the approach slabs. A propane gas-fired furnace is used to heat a mixture of propylene glycol and water. This antifreeze mixture circulates through a separate piping loop to the evaporators, heating the ammonia in the heat pipes. The bridge is tilted slightly on one end, so one end of the pipes is higher than the other. As the fluid boils, vapor rises in the heat pipes from the lower end to the higher and warms the bridge deck. As the vapor cools, it condenses and trickles back to the evaporators, where it is reheated (Fig. 10).

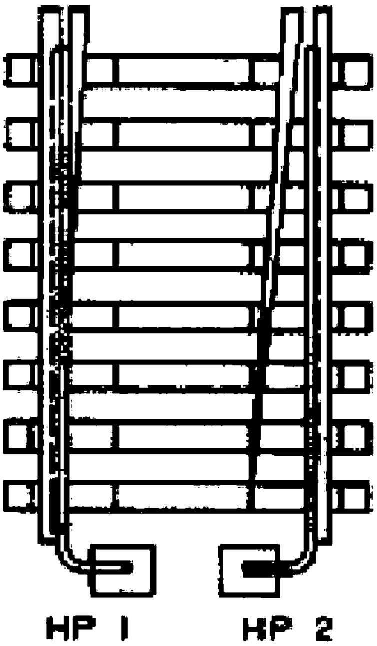


Fig. 9 Heat pipes for rail-point heating in winter.
Source: From Elsevier (see Ref. 11).

ENGINES AND THE AUTOMOTIVE INDUSTRY

Many sources of heat are available for low-temperature applications, including solar ponds, industrial waste, and geothermal energy. These sources of heat could be utilized for power generation. Many attempts to use this low-grade heat to produce electricity have been proposed. One of the attempts to integrate a turbine with a heat pipe was introduced in the Thermosyphon Rankine (TSR) engine for generating electricity. The heat pipe used in this application transfers heat from the evaporator through the turbine to the condenser. The schematic of the TSR engine is shown in Fig. 11. The working fluid located in the lower evaporator end of the pipe is evaporated and rises to the upper region with the application of the heat source. It is fed through the turbine, which in turn generates electricity.

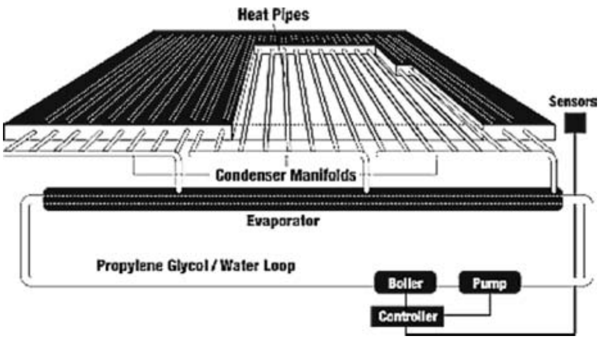


Fig. 10 Heat pipe transportation.
Source: From Geo-Heat Centre (see Ref. 13).

Heat–Inde

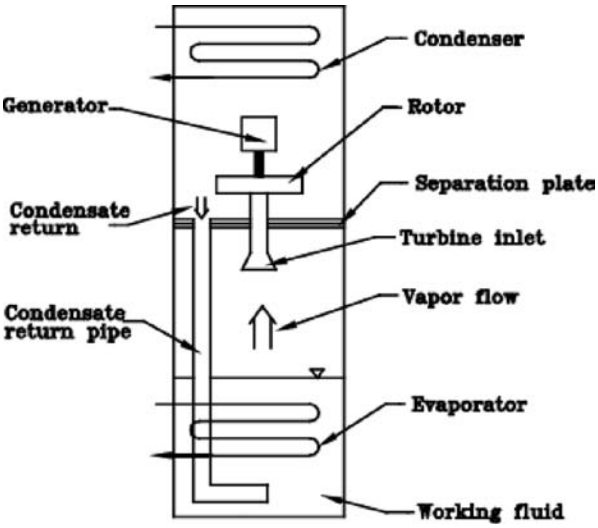


Fig. 11 Schematic of the TSR engine.
Source: From Elsevier (see Ref. 14).

This is an example of the conversion of kinematic energy to electrical energy.

OTHER HEAT PIPE APPLICATIONS

Cutting Tools

Cutting processes are being used in many factories. In the process, the energy supplied is converted to heat energy at the cutting zone. The heat generated in the tool and work piece is carried away by the working fluid. Three types of fluids[15] are commonly used: oil (with additives such as sulfur, chlorine, and phosphorus), emulsions, and synthetics. Although this cooling method shows some promise in increasing tool life, the exposure of the fluids to the working environment may cause significant contamination to the environment and increased health hazards to the workers. Currently, there is a strong worldwide trend toward minimizing the use of cutting fluids, as they have been shown to be a primary source of industrial pollution. For a cutting process to run without the use of fluids, alternative methods to remove the heat accumulated in the tool and the work piece had to be developed. Heat pipes provide an effective alternative solution to the problem, as shown in Fig. 12.

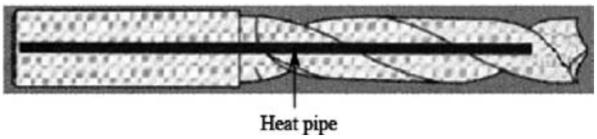


Fig. 12 Heat pipe cooling in drilling applications.
Source: From Elsevier Science (see Ref. 15).

CONCLUSION

Heat pipes have been used in many applications. This is because of their high thermal conductivity and their ability to maintain a constant temperature while heat flux varies. The ability to block heat flow when that is needed and the ability to dissipate heat from locations where cooling space is limited make the application of heat pipe technology indispensable.

Nowadays, gasoline poses a major expense in daily life. On hot summer days, while a car is parked, it will absorb a great deal of heating from solar radiation. The potential exists to use this heat in useful ways, rather than combating the heat by turning on the air conditioner. This is a situation in which established heat pipe technology can be applied effectively with little modification. If adopted, such application will also result in less gasoline consumption, thus helping keep the price of gasoline from rising even higher.

In medicine, heat pipes can also be introduced into back braces for use in people with back problems for which the application of constant heat affords some relief. After the acute phase of traumatic injury (e.g., a muscle tear or simple fracture), the application of heat using heat pipes will help reduce swelling and promote rapid healing.

Similarly, the application of heat pipe technology can make more efficient the application of constant low heat (presently achieved with cumbersome heat packs that have to be replaced every few hours), which can help relieve the aches and pains commonly seen in athletes and in members of the older generation.

ACKNOWLEDGMENTS

The author wishes to thank John W. Wang for his editorial advice and Dr. John Duffy for his support of my vision for this article.

REFERENCES

1. Amir, F. *Heat Pipe Applications, Heat Pipe Science And Technology*; Taylor & Francis: New York, NY. 1995; 41–55.
2. Volodymyr, B.; Stanislay, Z.; Jaroslav, V.; Franz, L. Elaboration of thermal control systems on heat pipes for microsatellites magion 4, 5 and BIRD. Appl. Therm. Eng. **2003**, *23*, 1109–1117.
3. Xiao, P.W.; Peter, J.; Aliakbar, A. Application of heat pipe exchangers to humidity control in air-conditioning systems. Appl. Therm. Eng. **1997**, *17*, 561–568.

4. Leonard, L.V. Review heat pipes in modern heat exchangers. Appl. Therm. Eng. **2005**, *25*, 1–19.
5. Feng, Y.; Xiugan, Y.; Guiping, L. Waste heat recovery using heat pipe heat exchanger for heating automobile using exhaust gas. Appl. Therm. Eng. **2003**, *23*, 367–372.
6. Kwang-Soo, K.; Myong-Hee, W.; Jong-Wook, K.; Byung-Joon, B. Heat pipe cooling technology for desktop PC CPU. Appl. Therm. Eng. **2003**, *23*, 1137–1144.
7. Hong, Z.; Jun, Z. Research, development and industrial application of heat pipe technology in China. Appl. Therm. Eng. **2003**, *23*, 1067–1083.
8. Voraos, N.; Kiatsiriroat, T.; Vorayos, N. Performance analysis of solar ethanol distillation. Renew. Energy **2006**, *31*, 2543–2554.
9. Mehmet, E. Thermal performance of a solar cooker integrated vacuum-tube collector with heat pipes containing different refrigerants. Sol. Energy **2004**, *76*, 751–757.
10. Srihajong, N.; Ruamrungsri, S.; Terdtoon, P.; Kamonpet, P. Heat pipe as a cooling mechanism in an aeroponic system. Appl. Therm. Eng. **2006**, *26*, 267–276.
11. Vasiliev, L.L. State-of-the-art on heat pipe technology in the former soviet union. Appl. Therm. Eng. **1998**, *18*, 507–551.
12. Levine, A.H.; Meade, J.C. Electrosurgery with Cooled Electrodes. U.S. Patent 7,074,219.
13. John, W.L. Pavement Snow Melting; GHC Bulletin, June 2000; 17–18. http://geoheat.oit.edu/bulletin/bull21-2/art4.pdf (accessed July, 2006).
14. Akbarzadeh, A.; Johnson, P.; Nguyen, T.; et al. Formulation and analysis of the heat pipe turbine for production of power from renewable sources. Appl. Therm. Eng. **2001**, *21*, 1551–1563.
15. Jen, T.C.; Gutierrez, G.; Eapen, S.; et al. Investigation of heat pipe cooling in drilling applications. Part I: preliminary numerical analysis and verification. Int. J. Mac. Tools Manufacture **2002**, *42*, 643–652.

Heat Pumps

Lu Aye
International Technologies Centre (IDTC), Department of Civil and Environmental Engineering, The University of Melbourne, Victoria, Australia

Abstract
Heat pumps capture heat energy from low-grade heat sources, such as ambient air and waste heat streams, and upgrade to a higher temperature level for useful applications. This entry includes a brief history, fundamentals, classifications, applications, and performance parameters of heat pumps. Working principles of the thermoelectric heat pump, the absorption heat pump, the gas compression heat pump, and the vapor compression heat pump are explained. Performance parameters of the heat pump systems are presented: coefficient of performance, energy efficiency ratio, primary energy ratio, and ambient energy fraction.

INTRODUCTION

This entry aims to provide foundation knowledge about heat pumps for both professional and nonprofessional readers. A brief history of heat pumps, basic terms used in the heat pump industry and research, and the fundamentals of heat pumps are presented. Working principles of the thermoelectric heat pump, the absorption heat pump, the gas compression heat pump, and the vapor compression heat pump are explained by using schematic diagrams. Performance parameters of the heat pump systems—coefficient of performance (COP), energy efficiency ratio (EER), primary energy ratio (PER), and ambient energy fraction (AEF)—are also discussed.

Major advances have been made in heat pump technologies over the past 30 years. Heat pumps in their various and diverse forms allow one to harness, in a cost-effective manner, solar energy that has already been stored in the atmosphere and biosphere. To make a fair systematic comparison with other systems that provide the same thermal output, AEF may be used. The concept and the method of estimating the AEF are discussed. The relationships among the AEF, the heating coefficient of performance (HCOP), and PER are also presented.

Brief History

Although most people are familiar with the common household refrigerator, the concept of using such a device to provide heating rather than cooling is less widely understood. The basic operational principle of this machine called heat pump, however, was laid down in thermodynamic terms by Lord Kelvin, the first professor of natural philosophy at the University of Glasgow, Scotland in the middle of the 19th century. Whereas the thermodynamic principle of the heat pump was found in 1852, it was not until 1855 that it was realized for producing heat by means of an open-cycle mechanical vapor recompression unit in Ebensee, Austria.[1] Much later, the closed vapor compression process was used for generating useful heat. After World War II, heat pump units for air conditioning homes and individual rooms became common. Now heat pump technology is well known as one of the energy conservation technologies.

Fundamentals

A heat pump is a thermodynamic system whose function is to heat at the required temperature with the aid of heat extracted from a source at lower temperature.[2] Heat pumps are devices designed to utilize low-temperature sources of energy that exist in atmospheric air, lake or river water, and in the earth. These sources are referred to as ambient energies. Heat pumps can also be operated using waste heat from commercial and industrial processes and thereby upgrading this to the required temperature level for some thermal process operations.[3]

The heat pump can be considered to be a heat engine in reverse.[4] A heat engine removes heat from a high-temperature source and discharges heat to a low-temperature sink, and in doing so can deliver work (see Fig. 1). Heat pumps capture heat energy from low-grade sources and deliver at a higher temperature. They require external work (W) to upgrade the heat absorbed (Q_L) (see Fig. 2). Note the heat energy flows from a low-temperature heat source to a higher application temperature. This is reverse to the natural flow direction of heat from a higher temperature to a lower temperature. For ideal conditions without heat losses the energy balance of a heat pump provides (Eq. 1):

$$Q_H = Q_L + W \tag{1}$$

Keywords: Heat pumps; Energy conservation; Heating; Energy recovery; Low-grade heat; Coefficient of performance; Energy efficiency ratio; Primary energy ratio; Ambient energy fraction.

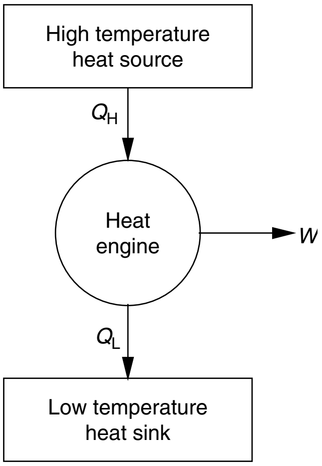


Fig. 1 Thermodynamic model of heat engine.

where Q_H, useful heat delivered; Q_L, heat absorbed; W, work input.

The useful heat delivered is always greater than the work input to the heat pump for ideal conditions.

The technical and economic performance of a heat pump is closely related to the characteristics of the heat source. Table 1 lists current commonly used heat sources. Ambient and exhaust air, soil, and ground water are practical heat sources for small heat pump systems, whereas sea/lake/river water, rock (geothermal), and waste water are used for large heat pump systems.[5]

The fundamental theories of heat pumps and refrigerators are the same. In engineering practice, a distinction is made between heat pumps and refrigerators. The principal purpose of the latter devices is to remove heat from a low-temperature heat source, and the purpose of the former device is to supply heat to a high-temperature sink. In many practical devices, the heating and cooling elements are interchangeable. The term reverse cycle air conditioner is frequently applied to heat pumps with interchangeable heating and cooling elements. In Japan and in the United States, reversible air-conditioning units are called heat pumps, and in Europe, the term heat pump is used for heating-only units.[1]

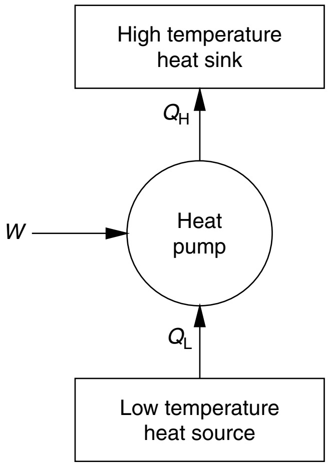


Fig. 2 Energy flow of a heat pump.

Table 1 Commonly used heat sources

Heat source	Temperature range (°C)
Ambient air	−10 to 15
Building exhaust air	15–25
Ground water	4–10
Lake water	0–10
River water	0–10
Sea water	3–8
Rock	0–5
Ground	0–10
Waste water and effluent	>10

Carnot Coefficient of Performance (Theoretical COP)

The heating performance of a heat pump is measured by HCOP. The HCOP is the ratio of the quantity of heat transferred to the high-temperature sink to the quantity of energy driving the heat pump (see Fig. 2).

$$\text{HCOP} = \frac{Q_H}{W} \tag{2}$$

In the Carnot ideal heat pump cycle, the heat is delivered isothermally at T_H and received isothermally at T_L. By using the laws of thermodynamics and definition of entropy, it can be shown that the Carnot heating coefficient of performance is given by Eq. 3.

$$\text{HCOP} = \frac{T_H}{T_H - T_L} = \frac{T_L}{T_H - T_L} + 1 \tag{3}$$

No practical heat pump constructed can have a better performance than this theoretical ideal HCOP. All that our practical heat pump cycles can do is struggle toward achieving this performance.[4]

Facts About Heating Relevant to Heat Pumps

It is argued that heat pumps are very energy efficient and, therefore, environmentally benign. The IEA heat pump centre provides the basic facts about heat supply and discusses the value of heat pumps.[6] The basic facts about heating explained are

- Direct combustion to generate heat is never the most efficient use of fuel.
- Heat pumps are more efficient because they use renewable energy in the form of low-temperature heat.

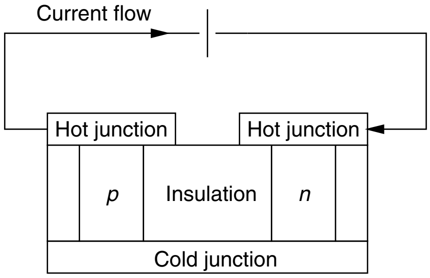


Fig. 3 Schematic of thermoelectric heat pump.

- If the fuel used by conventional boilers were redirected to supply power for electric heat pumps, about 35%–50% less fuel would be needed, resulting in 35%–50% less emissions.
- Around 50% savings are made when electric heat pumps are driven by combined heat and power (CHP) or cogeneration systems.
- Whether fossil fuel, nuclear energy, or renewable power is used to generate electricity, electric heat pumps make far better use of these resources than do resistance heaters.
- The fuel consumption and, consequently, the emissions rate of an absorption or gas-engine heat pump are about 35%–50% less than those of a conventional boiler.

Because heat pumps consume less primary energy than conventional heating systems, they are an important technology for reducing the unwanted air pollutants that harm the human environment, such as respirable particulate matters (PM), carbon monoxide (CO), nitrogen oxides (NO_x), sulphur dioxide (SO_2), and carbon dioxide (CO_2).

HEAT PUMP TYPES

Four major types of heat pumps may be identified in the literature.[7] These are

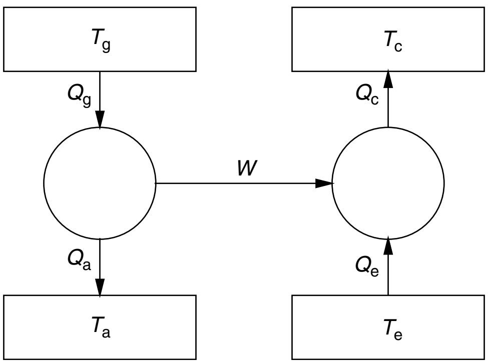


Fig. 4 Thermodynamic model of absorption heat pump.

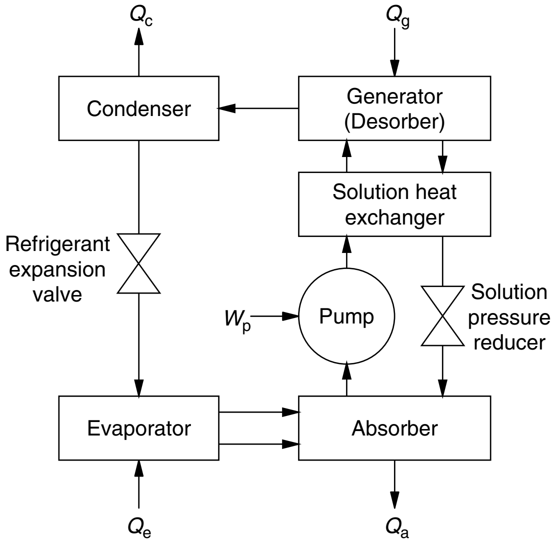


Fig. 5 Simple absorption cycle.

- The thermoelectric heat pump
- The absorption heat pump
- The gas compression heat pump
- The vapor compression heat pump

The Thermoelectric Heat Pump

The thermoelectric heat pump uses the Peltier effect. Two materials with different thermoelectric properties (p type and n type) are arranged to form hot and cold junctions, and are connected in series with a direct-current voltage source (Fig. 3). These devices do not have moving parts, and the cooling and heating elements may be reversed simply by changing the direction of the current.

The Absorption Heat Pump

All absorption cycles include at least three thermal exchanges with their surroundings (i.e., energy exchange at three different temperatures). For the system to operate, the generator temperature (T_g) must be greater than the condenser temperature (T_c) and the absorber temperature (T_a), which in turn must be greater than the evaporator temperature (T_e). An absorption cycle is a heat-activated thermal cycle. The absorption heat pump can be considered as a heat engine driving a heat pump (see Fig. 4).

The components of a simple single-effect absorption system are

- Generator (desorber)
- Condenser
- Refrigerant expansion valve
- Evaporator

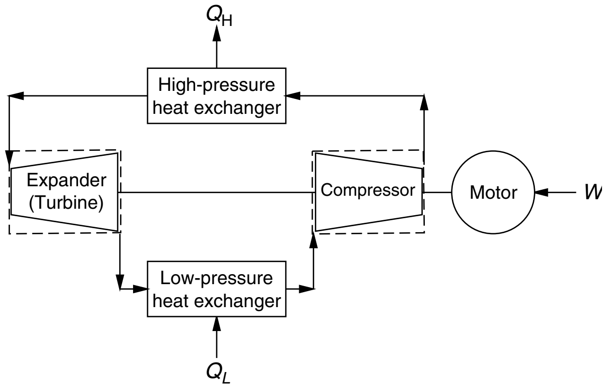


Fig. 6 Schematic of gas compression heat pump.

- Absorber
- Pump
- Solution heat exchanger
- Solution pressure reducer
- Connecting pipes

The absorption system utilizes a sorbent–refrigerant pair. The widely used pairs are Lithium-bromide and water, and water and ammonia. A schematic of the continuous absorption heat pump system is shown in Fig. 5. An absorption cycle is a heat-activated thermal cycle. The major exchanges with its surroundings are thermal energy. A small amount of mechanical work is required at the pump to transport the working fluids.

A high-temperature heat source supplies heat to the generator to drive off the refrigerant vapor from solution. The condenser receives the refrigerant vapor from the generator. In the condenser, the refrigerant changes phase from vapor to liquid, and heat is rejected at a medium temperature. The refrigerant expansion valve reduces the pressure of the refrigerant from the condensing pressure to the evaporating pressure. In the evaporator, the liquid refrigerant collects the low-temperature heat and changes to vapor phase. In other words, the working fluid (refrigerant) provides the cooling effect at the evaporator, and the vaporized refrigerant enters the absorber. The strong-concentration sorbent absorbs the refrigerant vapor and becomes a weak solution. Note the heat rejected at the absorber. The low-pressure weak solution (refrigerant and sorbent) is pumped to the solution heat exchanger to exchange heat with the strong-concentration sorbent. The solution pressure reducer accepts the high-pressure strong solution from the generator and delivers low-pressure strong solution to the absorber. The generator receives the high-pressure weak solution from the solution heat exchanger, and separates the refrigerant and sorbent.

The Gas Compression Heat Pump

Gas compression heat pumps use a gas as the working fluid. Phase changes do not occur during the cycle of operation. Major components of a gas compression heat pump system are:

- Gas compressor
- High-pressure heat exchanger
- Gas expander (or turbine)
- Drive motor
- Low-pressure heat exchanger

A schematic of a gas compression heat pump is shown in Fig. 6. The gas is drawn into the compressor and heated. The heated high-pressure gas delivers heat at the high-pressure heat exchanger. At the expander (turbine), high-pressure gas does the expansion work that helps drive the compressor. In general, the external work input to the compressor is provided by a drive motor.

The Vapor Compression Heat Pump

Almost all heat pumps currently in operation are based on either a vapor compression or an absorption cycle.[8] Heat pumps and refrigerators of the vapor compression type utilize a working fluid that undergoes phase changes during operation. The phase changes occur principally during the heat collection and rejection stages of the cycles. The phase-changing processes occurring in the vapor compression heat pumps are accompanied by extremely favorable heat transfer conditions. Boiling agitation in evaporators and dropwise condensation in condensers make the heat transfer rates very high. To carry out evaporation and condensation, any practical heat pump uses four separate components, as illustrated in Fig. 7:

- Compressor
- Condenser
- Expansion device
- Evaporator

The compressor is the heart of the vapor compression heat pump system. By using mechanical power, the compressor increases the pressure of the working fluid vapor received from the evaporator and delivers it to the condenser. The working fluid condenses and provides useful heat at the condenser. The pressure of the liquid working fluid coming out of the condenser is reduced to the evaporating pressure by an expansion device. The low-grade heat is collected at the evaporator, and the phase of the working fluid changes from liquid to vapor. The working fluid enters the compressor in vapor form to repeat the cyclic flow.[9]

Compressors used in practical units are generally of the positive displacement type: either reciprocating or rotary vanes. In practice, nearly all vapor compression heat pumps use thermostatic expansion valves (T-X valves) because of the availability of the valve to handle a wide range of operating conditions and because of the fact that

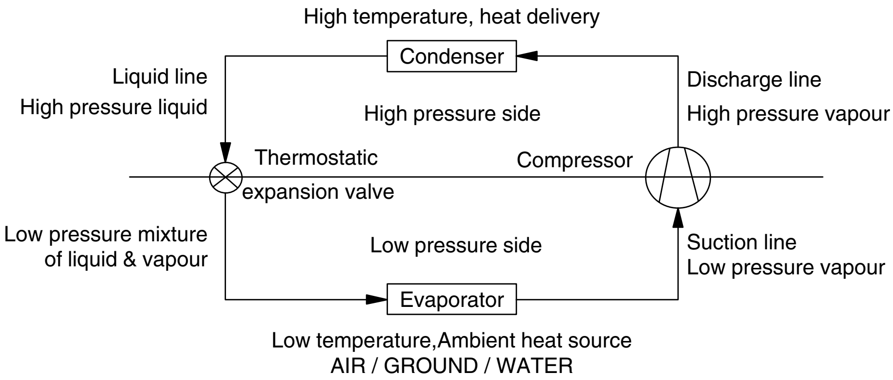


Fig. 7 Schematic diagram of vapor compression heat pump.

the pressure reduction is essentially isenthalpic. These valves, which are relatively low-cost devices, control the liquid working fluid flow to the evaporator by sensing the superheated condition of the working fluid leaving the evaporator. Such control ensures that nearly all the available evaporator surface is covered with a forced convection nucleate boiling film with consequential excellent heat transfer characteristics in the evaporation process.

Reversible circuit heat pumps for building cooling and heating are equipped with a four-way exchange valve that can be operated automatically or manually. This valve enables the normal evaporator to become the condenser and the normal condenser to become the evaporator so that the machine can be made to operate in reverse fashion. This arrangement is often incorrectly called a reverse-cycle operation because of the end effect achieved—i.e., the interchange of roles for the cooling and heating parts of the circuit. It would be better to call this a reverse-flow operation, as the basic thermodynamic cycle remains the same for each mode of operation.

In general, vapor compression heat pumps are classified in many ways. One method commonly used distinguishes among air, water, and ground sources of low-temperature energy and also between the high-temperature energy delivery medium (i.e., air or water). In such a classification, the most common types of heat pump units would include air to air units and water to water units. Solar energy could also be captured at the evaporator as a heat source. These types of heat pumps are called solar-assisted and sola-boosted heat pumps.[10]

Changes in condensing and evaporating temperatures affect the work done at the compressor, and, consequently the COP. Decreasing the condenser or increasing the evaporator temperatures will decrease the compressor work and increase the COP. Fig. 8 shows the effect of condensing temperature on COP for a fixed evaporator temperature. The effect of the evaporating temperature on the COP for a fixed condenser load temperature is shown in Fig. 9. For simplicity, these curves have been obtained assuming operation on the ideal heat pump cycle—i.e., the Carnot cycle of two isothermal and two isentropic (reversible adiabatic) processes not allowing any practical deviations. Although the COP values for real machines will be substantially lower, the trend will be the same.

PERFORMANCE PARAMETERS

This section defines and explains some performance parameters that are used to compare various heat pumps

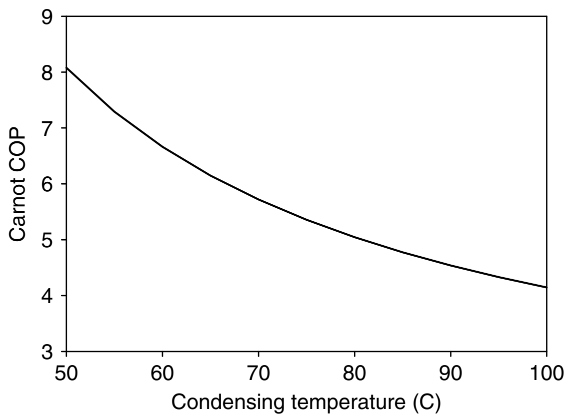


Fig. 8 Effect of condensing temperature on ideal COP (evaporating temperature = 10°C).

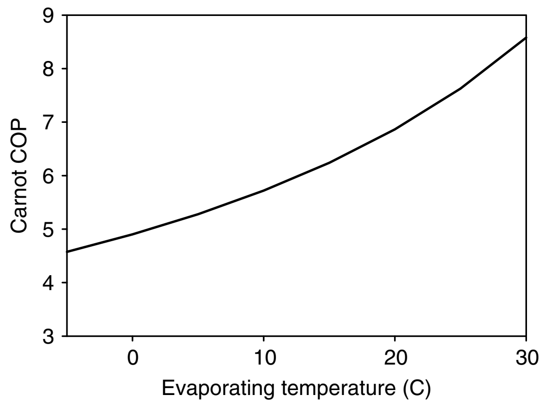


Fig. 9 Effect of evaporating temperature on ideal COP (condensing temperature = 70°C).

Heat–Inde

for a particular application, and also presents the relationships among them.

Coefficient of Performance and Energy Efficiency Ratio

System coefficient of performance (SCOP) is defined as output heating capacity per unit of power input to the system (see Eq. 4).

$$\mathrm{SCOP} = \frac{\text{Heating capacity(W)}}{\text{Power input(W)}} \tag{4}$$

System coefficient of performance is a dimensionless number that measures the performance of a heat pump. If the heating capacity is expressed in units other than W, it is called the heating energy efficiency ratio (HEER). Heat pumps sold in the United States are often stated in terms of EER (see Eq. 5).

$$\mathrm{HEER} = \frac{\text{Heating capacity(Btu/h)}}{\text{Power input(W)}} \tag{5}$$

In the equation described, EER has a unit of Btu/h per W. It should be noted that 1 Btu/h = 0.2928 W. Therefore, the relationship between EER and COP can be expressed as in Eq. 6.

$$\mathrm{COP} = \mathrm{EER} \times 0.2928 \tag{6}$$

Primary Energy Ratio

The HCOP provides a measure of the usefulness of the heat pump system in producing large amounts of heat from a small amount of work. It does not express the fact that energy available as work is normally more useful than energy available as heat. To assess different heat pump systems using compressor drives from different fuel or energy sources, the PER is applied. The PER takes into account not only the heat pump COP, but also the efficiency of conversion of the primary fuel into the work that drives the compressor. Primary energy ratio is defined as in Eq. 7.[4] This also can be expressed as in Eq. 8,

$$\mathrm{PER} = \frac{\text{Useful heat delivered by heat pump}}{\text{Primary energy consumed}} \tag{7}$$

$$\mathrm{PER} = \frac{Q_H}{E_{pe}} = \frac{Q_H}{W} \times \frac{W}{E_{pe}} = \mathrm{HCOP} \times \eta_{pp} \tag{8}$$

where Q_H = load, E_{pe} = primary energy used by the heat pump system, and η_{pp} = power plant efficiency.

The drive energy of heat pumps is most commonly electricity. "Ideally, a heat pump where free work is available should be contemplated, e.g., wind or water power."[4] Consider an electric heat pump powered by a

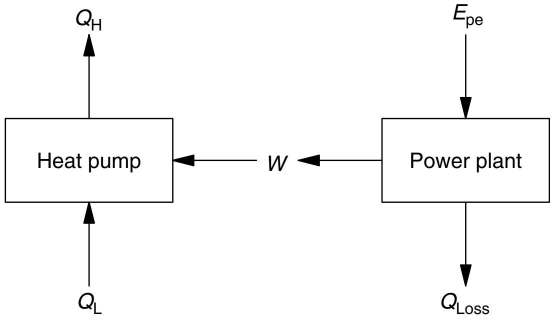


Fig. 10 Energy flow of a conventional electric heat pump.

conventional power plant fuelled by a nonrenewable energy source (see Fig. 10). The power-plant efficiency, η_{pp}, is up to 58% for oil- or gas-fired combined-cycle power plants currently available on the market. The PER is equal to the HCOP for direct power generation from renewable ambient energy sources, such as solar and wind.[1] This concept is illustrated in Fig. 11.

The amount of renewable or ambient energy spent (used up) to produce work for the heat pump is equal to the amount of work (i.e., $E_{re} = W$)—for this case, by definition $\eta_{pp} = 1$.[1] It should be noted that unlike the losses in a fossil-fuel power plant, the unused ambient energy passing through the renewable power plant is still in the form of ambient energy, and it is available to be used.

Ambient Energy Fraction

To make a fair systematic comparison of heat pumps and other systems that provide the same heat output, the term AEF was developed by Aye et al.[11] The term solar fraction is widely understood, accepted, and used in the solar energy field. It is defined as the fractional reduction of purchased energy when a solar energy system is used.[12] It is the fraction of the load contributed by the solar energy, which can be calculated by Eq. 9,

$$f = \frac{L - E}{L} = \frac{L_s}{L} \tag{9}$$

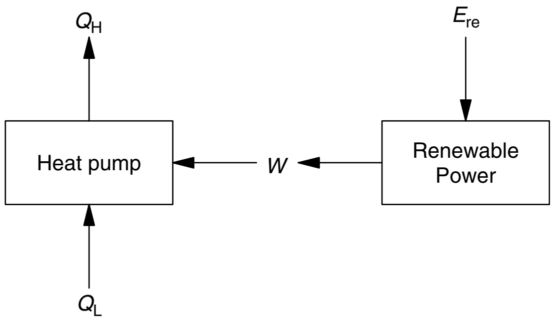


Fig. 11 Energy flow of a renewable electric heat pump.

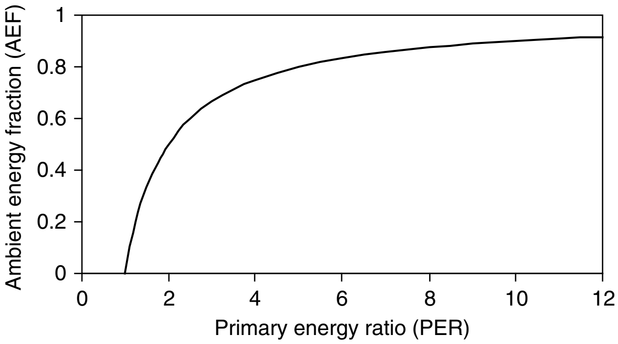


Fig. 12 The ambient energy fraction vs the primary energy ratio.

where L, the load; E, the auxiliary energy supplied to the solar energy system; and L_s, the solar energy delivered.

Similar to the solar fraction of a solar system, the term AEF of a heat pump can be defined as the fraction of the load contributed by the ambient energy, which may be calculated as in Eq. 10.

$$\mathrm{AEF} = \frac{Q_H - E_{pe}}{Q_H} = 1 - \frac{E_{pe}}{Q_H} = 1 - \frac{1}{\mathrm{PER}} = 1 - \frac{1}{\mathrm{HCOP} \times \eta_{pp}} \tag{10}$$

Fig. 12 illustrates the relationship between the AEF and the PER.

Table 2 shows the AEFs of a heat pump, which has an HCOP of 3.5, for various electric power plants. Table 2 clearly shows that for the same heat pump, the AEF may vary from 1 to 71%, depending on how electricity is generated for driving the heat pump compressor. The AEF for the renewable electricity-driven heat pump is the highest (71%). The renewable energy used is only 29% (i.e., 100%–71%) of the total thermal load. The energy use of the brown-coal electricity-driven heat pump is 99% of the total thermal load (i.e., 0.99 MJ of brown-coal energy is required for 1 MJ of thermal load). Table 3 shows the effect of heat pump COP on the AEF of a heat pump powered by a renewable source.

HEAT PUMP APPLICATIONS

Heat pumps have been used for domestic, commercial, and industrial applications.

Domestic applications are:

- Provision of space heating
- Provision of hot water
- Swimming-pool heating

Table 2 Ambient energy fractions of a heat pump powered by typical power plants

Power plant	Brown coal	Nuclear	Black coal	Gas-fired combined-cycle	Renewable
Power plant efficiency, η_{pp}	0.29[a]	0.33[b]	0.35[a]	0.58[b]	1.00
HCOP of heat pump	3.50	3.50	3.50	3.50	3.50
Primary energy ratio, PER	1.02	1.16	1.23	2.03	3.50
Ambient energy fraction, AEF	0.01	0.13	0.18	0.51	0.71

[a]Typical Australian data

Source: From DPIE, Commonwealth Department of Primary Industries and Energy/Australian Cogeneration Association (see Ref. 13).

[b]Typical European data.

Source: From International Institute of Refrigeration. (see Ref. 1).

Table 3 Effect of HCOP on AEF of a heat pump powered by renewable source

HCOP of heat pump	1.50	2.50	3.50	4.50
Renewable power plant efficiency, η_{pp}	1.00	1.00	1.00	1.00
Primary energy ratio, PER	1.50	2.50	3.50	4.50
Ambient energy fraction, AEF	0.33	0.60	0.71	0.78

Commercial and industrial applications are

- Space heating
- Water heating
- Swimming-pool heating
- Drying and dehumidification
- Evaporation and boiling
- Desalination

CONCLUSION

The fundamentals of heat pumps were presented together with the working principles of the thermoelectric heat pump, the absorption heat pump, the gas compression heat pump, and the vapor compression heat pump. It should be noted that vapor compression heat pumps driven by electricity dominate the current market. Heat pumps are very energy efficient and, therefore, environmentally benign compared with other available heating technologies. The technical and economic performance of a heat pump is closely related to the characteristics of the heat source.

Various performance parameters are available for comparisons of heat pumps; HCOP and energy efficiency ratio are the most widely used. Primary energy ratio and AEF can be used to compare various heat pump systems systematically and fairly. The AEF of an electric heat pump depends on the HCOP and the power plant efficiency based on the primary energy used. The AEF is highly dependent on the type of power plant used to generate the electricity that drives the heat pump compressor. Heat pumps driven by renewable electricity offer the possibility of reducing energy consumption significantly. In the future, the AEF may be used widely as a performance parameter of heat pumps, because energy-resource issues are becoming more important.

ACKNOWLEDGMENTS

I wish to thank Emeritus Professor William W.S. Charters, former Dean of Engineering at the University of Melbourne, who introduced me to heat pump technology and for supporting my career and research in heat pumps.

REFERENCES

1. Halozan, H.; Rieberer, R. Energy-efficient heating and cooling systems for buildings. IIR Bull **2004**, *LXXXI* (2004-6), 6–22.
2. Gac, A. et al. *Guide for the Design and Operation of Average and Large Capacity Electric Heat Pumps*; International Institute of Refrigeration: Paris, 1988.
3. McMullen, J.T.; Morgan, R. *Heat Pumps*; Adam Hilger Ltd.; Bristol, U.K., 1981.
4. Reay, D.A.; Macmichael, D.B.A. *Heat Pumps Design and Applications: A Practical Handbook for Plant Managers, Engineers, Architects and Designers*; Pergamon Press: Oxford, U.K., 1979.
5. IEAHPC (see International Energy Agency Heat Pump Centre). Heat Sources. Available at: www.heatpumpcentre.org/About_heat_pumps/Heat_sources.asp (accessed September 2005).
6. IEAHPC (see International Energy Agency Heat Pump Centre). How Heat Pumps Achieve Energy Savings and CO_2 Emissions Reduction: An Introduction. Available at: www.heatpumpcentre.org/About_heat_pumps/Energy_and_CO2.asp (accessed September 2005).
7. Taylor L.E. The design, construction, and evaluation of a solar boosted heat pump, Master of Engineering Science thesis, Department of Mechanical Engineering, The University of Melbourne: Melbourne, Australia, 1978.
8. IEAHPC (see International Energy Agency Heat Pump Centre). Heat Pump Technology. Available at: www.heatpumpcentre.org/About_heat_pumps/HP_technology.asp (accessed September 2005).
9. Aye, L.; Charters, W.W.S. Electrical and Engine Driven Heat Pumps for Effective Utilisation of Renewable Energy Resources. Appl. Therm. Eng. **2003**, *23* (10), 1295–1300.
10. Aye L.; Charters W.W.S.; Chaichana C. Chapter 7: Solar boosted heat pump (SBHP). In *Solar Energy: New Research*, Hough, T.P., Ed., ISBN: 1-59454-630-4; Nova Science Publishers, Inc.; New York, 2006; 239–254.
11. Aye L.; Fuller R.J.; Charters W.W.S. Ambient energy fraction of a heat pump. In *Solar 2005, Renewable Energy for a Sustainable Future: A Challenge for a Post Carbon World*, Proceedings of the 43rd Annual Conference of the Australian and New Zealand Solar Energy Society, Dunedin, New Zealand, 28–30 November 2005.
12. Duffie, J.A.; Beckman, W.A. *Solar Engineering of Thermal Processes*, 2nd Ed.; Wiley Interscience publication: New York, 1991.
13. Taylor, M.; Labson, S. *Profiting from Cogeneration*; DPIE, Commonwealth Department of Primary Industries and Energy/Australian Cogeneration Association: Canberra, Australia, 1997.

Heat Transfer

Yunus A. Cengel
Department of Mechanical Engineering, University of Nevada, Reno, Nevada, U.S.A.

Abstract
Heat can be transferred in three different modes: conduction, convection, and radiation. All modes of heat transfer require the existence of a temperature difference, and all modes are from the high-temperature medium to the lower-temperature medium. Conduction is the transfer of energy from the more energetic particles of a substance to the adjacent, less energetic ones as a result of interactions between the particles. Convection is the mode of heat transfer between a solid surface and the adjacent liquid or gas that is in motion, and it involves the combined effects of conduction and fluid motion. Radiation is the energy emitted by matter in the form of electromagnetic waves (or photons) as a result of the changes in the electronic configurations of the atoms or molecules. In this section, we present an overview of the three basic mechanisms of heat transfer and discuss thermal conductivity after giving some historical background.

INTRODUCTION

Heat or thermal energy is the form of energy that can be transferred to or from a system by virtue of a temperature difference. An energy interaction is heat transfer only if it takes place because of a temperature difference. Therefore, there cannot be any heat transfer between two systems that are at the same temperature.

The science of thermodynamics deals with the amount of heat transfer as a system undergoes a process from one equilibrium state to another and makes no reference to how long the process will take. But in engineering, we are often interested in the rate of heat transfer—the amount of heat transfer per unit time—which is the topic of the science of heat transfer.

Heat can be transferred in three different modes: conduction, convection, and radiation. All modes of heat transfer require the existence of a temperature difference and all modes are from the high-temperature medium to the lower-temperature medium. Conduction is the transfer of energy from the more energetic particles of a substance to the adjacent, less energetic ones as a result of interactions between the particles. Convection is the mode of heat transfer between a solid surface and the adjacent liquid or gas that is in motion, and it involves the combined effects of conduction and fluid motion. Radiation is the energy emitted by matter in the form of electromagnetic waves (or photons) as a result of the changes in the electronic configurations of the atoms or molecules. In this section, we present an overview of the three basic mechanisms of heat transfer and discuss thermal conductivity after giving some historical background.

HISTOPRICAL BACKGROUND

Heat has always been perceived to be something that produces in us a sensation of warmth and one would think that the nature of heat is one of the first things understood by mankind. However, it was only in the middle of the nineteenth century that we had a true physical understanding of the nature of heat, thanks to the development at that time of the kinetic theory, which treats molecules as tiny balls that are in motion and thus possess kinetic energy. Heat is defined as the energy associated with the random motion of atoms and molecules. Although it was suggested in the eighteenth and early nineteenth centuries that heat is the manifestation of motion at the molecular level (called the live force), the prevailing view of heat until the middle of the nineteenth century was based on the caloric theory proposed by the French chemist Antoine Lavoisier (1744–1794) in 1789. The caloric theory asserts that heat is a fluid-like substance called the caloric that is a massless, colorless, odorless, and tasteless substance that can be poured from one body into another. When caloric was added to a body, its temperature increased; and when caloric was removed from a body, its temperature decreased. When a body could not contain any more caloric, much the same way as when a glass of water could not dissolve any more salt or sugar, the body was said to be saturated with caloric. This interpretation gave rise to the terms saturated liquid and saturated vapor that are still in use today.

Keywords: Heat transfer; Thermal conductivity; Conduction; Convection; Radiation.

The caloric theory came under attack soon after its introduction. It maintained that heat is a substance that cannot be created or destroyed. Yet it was known that heat can be generated indefinitely by rubbing one's hands together or rubbing two pieces of wood together. In 1798, the American Benjamin Thompson (Count Rumford) (1754–1814) showed in his papers that heat can be generated continuously through friction. The validity of the caloric theory was also challenged by several others, but it was the careful experiments of the Englishman James P. Joule (1818–1889) published in 1843 that finally convinced the skeptics that heat was not a substance after all, thus putting the caloric theory to rest. Although the caloric theory was totally abandoned in the middle of the nineteenth century, it contributed greatly to the development of thermodynamics and heat transfer.

CONDUCTION

Conduction is the transfer of energy from the more energetic particles of a substance to the adjacent less energetic ones as a result of interactions between the particles. Conduction can take place in solids, liquids, or gases. In gases and liquids, conduction is due to the collisions and diffusion of the molecules during their random motion. In solids, it is due to the combination of vibrations of the molecules in a lattice and the energy transport by free electrons. A cold canned drink in a warm room, for example, eventually warms up to the room temperature as a result of heat transfer from the room to the drink through the aluminum can by conduction. A detailed study of these modes is given in Refs. 1–4.

The rate of heat conduction through a medium depends on the geometry of the medium, its thickness, and the material of the medium, as well as the temperature difference across the medium. We know that wrapping a hot-water tank with glass wool (an insulating material) reduces the rate of heat loss from the tank. The thicker the insulation is, the smaller the heat loss will be. We also know that a hot-water tank will lose heat at a higher rate when the temperature of the room housing the tank is lowered. Further, the larger the tank is, the larger the surface area and thus the rate of heat loss will be.

Consider steady heat conduction through a large plane wall of thickness $\Delta x = L$ and area A, as shown in Fig. 1. The temperature difference across the wall is $\Delta T = T_2 - T_1$. Experiments have shown that the rate of heat transfer $\dot{Q}$ through the wall is doubled when the temperature difference ΔT across the wall or the area A normal to the direction of heat transfer is doubled, but it is halved when the wall thickness L is doubled. Thus, we conclude that the rate of heat conduction through a plane layer is proportional to the temperature difference across the layer and the heat transfer area but is inversely proportional to the thickness of the layer. That is,

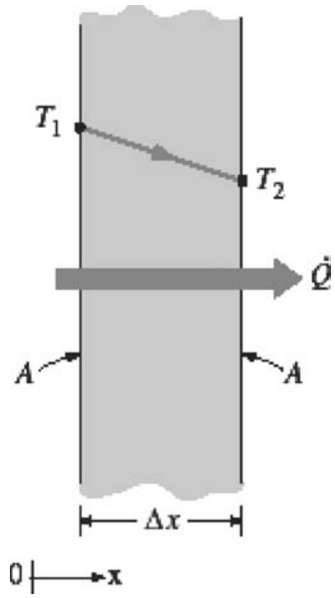


Fig. 1 Heat conduction through a large plane wall of thickness Δx and area A.

$$\text{Rate of heat conduction } \alpha \frac{\text{(Area)(Temperature difference)}}{\text{Thickness}}$$

or

$$\dot{Q} = kA\frac{T_1 - T_2}{\Delta x} = -kA\frac{\Delta T}{\Delta x} \quad (1)$$

where the constant of proportionality k is the thermal conductivity of the material, which is a measure of the ability of a material to conduct heat. The limiting case of $\Delta x \to 0$ reduces to the differential form:

$$\dot{Q} = -kA\frac{\mathrm{d}T}{\mathrm{d}x} \quad (2)$$

which is called Fourier's law of heat conduction after J. Fourier who expressed it first in his heat transfer text in 1822. Here $\mathrm{d}T/\mathrm{d}x$ is the temperature gradient, which is the slope of the temperature curve on a $T-x$ diagram (the rate of change of T with x), at location x. This relation indicates that the rate of heat conduction in a direction is proportional to the temperature gradient in that direction. Heat is conducted in the direction of decreasing temperature and the temperature gradient becomes negative when temperature decreases with increasing x. The negative sign in Eqs. 1 and 2 ensures that heat transfer in the positive x direction is a positive quantity.

The heat transfer area A is always normal to the direction of heat transfer. For heat loss through a 5-m-long, 3-m-high, and 25-cm-thick wall, for example, the heat transfer area is $A = 15\ \mathrm{m}^2$. Note that the thickness of the wall has no effect on A (Fig. 2).

Heat–Inde

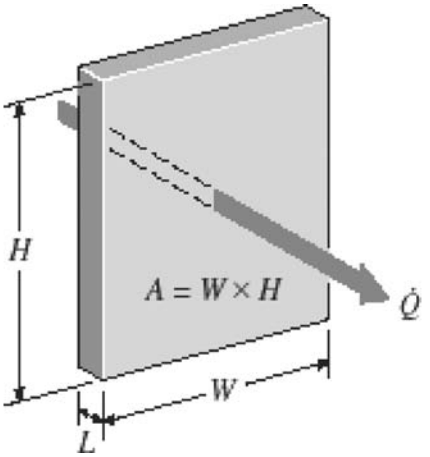


Fig. 2 In heat conduction analysis, A represents the area normal to the direction of heat transfer.

Example 1. The Cost of Heat Loss through a Roof.

The roof of an electrically heated home is 6 m long, 8 m wide, and 0.25 m thick, and is made of a flat layer of concrete whose thermal conductivity is $k = 0.8$ W/m K (Fig. 3). The temperatures of the inner and the outer surfaces of the roof one night are measured to be 15 and 4°C, respectively, for a period of 10 h. Determine the rate of heat loss through the roof that night and the cost of that heat loss to the home owner if the cost of electricity is $0.08/kWh.

Solution. Noting that heat transfer through the roof is by conduction and the area of the roof is $A = 6\text{ m} \times 8\text{ m} = 48\text{ m}^2$, the steady rate of heat transfer through the roof is determined to be

$$\begin{aligned}\dot{Q} &= kA\frac{T_1 - T_2}{L}\\ &= (0.8\text{ W/m K})(48\text{ m}^2)\frac{(15 - 4)^\circ\text{C}}{0.25\text{ m}} = 1690\text{ W}\\ &= 1.69\text{ kW}\end{aligned}$$

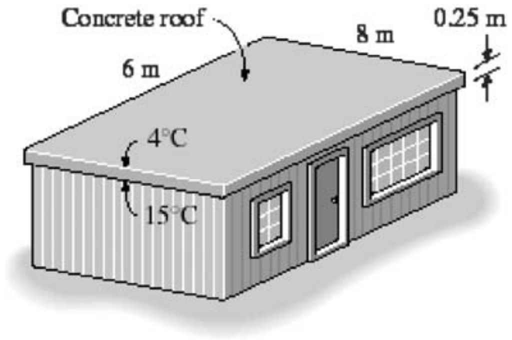


Fig. 3 Schematic for Example 1.

The amount of heat lost through the roof during a 10-hour period and its cost are

$$Q = \dot{Q}\Delta t = (1.69\text{ kW})(10\text{ h}) = 16.9\text{ kWh}$$

$$\begin{aligned}\text{Cost} &= (\text{Amount of energy})(\text{Unit cost of energy})\\ &= (16.9\text{ kWh})(\$0.08/\text{kWh}) = \$1.35\end{aligned}$$

Note that the total heating cost of the house is much larger than $1.35 because the heat losses through the walls are not considered. Also, for temperature differences, the units °C and K are identical.

Thermal Conductivity

Different materials store heat differently, and is defined as the property specific heat c_p is defined as a measure of a material's ability to store thermal energy. For example, $c_p = 4.18$ kJ/kg K for water and it is 0.45 kJ/kg K for iron at room temperature, which indicates that water can store almost 10 times the energy that iron can per unit mass. Likewise, the thermal conductivity k is a measure of a material's ability to conduct heat. For example, $k = 0.608$ W/m K for water and $k = 80.2$ W/m K for iron at room temperature, which indicates that iron conducts heat more than 100 times faster than water can. Thus we say that water is a poor heat conductor relative to iron, although water is an excellent medium to store thermal energy.

Eq. 2 for the rate of conduction heat transfer under steady conditions can also be viewed as the defining equation for thermal conductivity. Thus the thermal conductivity of a material can be defined as the rate of heat transfer through a unit thickness of the material per unit area per unit temperature difference. The thermal conductivity of a material is a measure of the ability of the material to conduct heat. A high value for thermal conductivity indicates that the material is a good heat conductor; and a low value indicates that the material is a poor heat conductor or an insulator. The thermal conductivities of some common materials at room temperature are given in Table 1. The thermal conductivity of pure copper at room temperature is $k = 401$ W/m K, which indicates that a 1-m-thick copper wall will conduct heat at a rate of 401 W through an area of 1 m^2 per °C or K temperature difference across the wall. Note that materials such as copper and silver that are good electric conductors are also good heat conductors and have high values of thermal conductivity. Materials such as rubber, wood, and Styrofoam are poor conductors of heat and have low conductivity values.

A layer of material of known thickness and area can be heated from one side by an electric resistance heater of known output. If the outer surfaces of the heater are well insulated, all the heat generated by the resistance heater

Heat–Inde

Table 1 The thermal conductivities of some materials at room temperature

Material	**k, W/m K[a]**
Diamond	2300
Silver	429
Copper	401
Gold	317
Aluminum	237
Iron	80.2
Mercury (*l*)	8.54
Glass	0.78
Brick	0.72
Water (*l*)	0.613
Human skin	0.37
Wood (oak)	0.17
Helium (g)	0.152
Soft rubber	0.13
Glass fiber	0.043
Air (g)	0.026
Urethane, rigid foam	0.026

[a]Multiply by 0.5778 to convert to Btu/h ft R.

will be transferred through the material whose conductivity is to be determined. Then, measuring the two surface temperatures of the material when steady heat transfer is established and substituting them into Eq. 2 together with other known quantities give the thermal conductivity (Fig. 4).

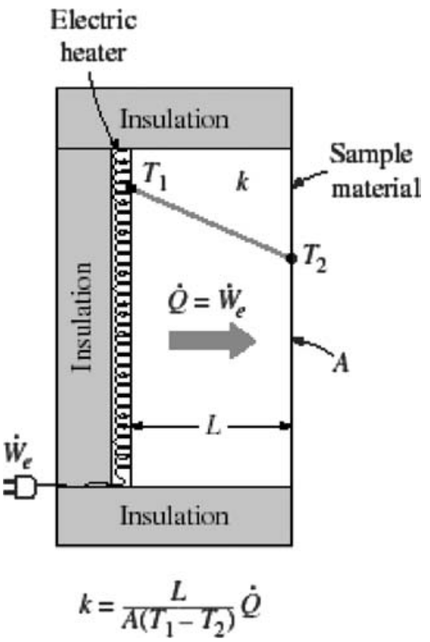


Fig. 4 A simple experimental setup to determine the thermal conductivity of a material.

The thermal conductivities of materials vary over a wide range, as shown in Fig. 5. The thermal conductivities of gases such as air vary by a factor of 10^4 from those of pure metals such as copper. Note that pure crystals and metals have the highest thermal conductivities and gases and insulating materials the lowest.

Temperature is a measure of the kinetic energies of the particles such as the molecules or atoms of a substance. In a liquid or gas, the kinetic energy of the molecules is due to their random translational motion as well as their vibrational and rotational motions. When two molecules possessing different kinetic energies collide, part of the kinetic energy of the more energetic (higher-temperature) molecule is transferred to the less energetic (lower-temperature) molecule, much the same as when two elastic balls of the same mass at different velocities collide—part of the kinetic energy of the faster ball is transferred to the slower one. The higher the temperature, the faster the molecules move, higher the number of such collisions, and the better the heat transfer.

The kinetic theory of gases predicts and the experiments confirm that the thermal conductivity of gases is proportional to the square root of the absolute temperature T and inversely proportional to the square root of the molar mass M. Therefore, the thermal conductivity of a gas increases with increasing temperature and decreasing molar mass. So it is not surprising that the thermal conductivity of helium ($M=4$) is much higher than those of air ($M=29$) and argon ($M=40$).

The mechanism of heat conduction in a liquid is complicated by the fact that the molecules are more closely spaced and they exert a stronger intermolecular force field. The thermal conductivities of liquids usually lie between those of solids and gases. The thermal conductivity of a substance is normally highest in the solid phase and lowest in the gas phase. Unlike gases, the thermal conductivities of most liquids decrease with increasing temperature, with water being a notable exception. Like gases, the conductivity of liquids decreases with increasing molar mass. Liquid metals such as mercury and sodium have high thermal conductivities and are very suitable for use in applications where a high heat transfer rate to a liquid is desired, as in nuclear power plants.

In solids, heat conduction is due to two effects: the lattice vibrational waves induced by the vibrational motions of the molecules positioned at relatively fixed positions in a periodic manner, called a lattice, and the energy transported via the free flow of electrons in the solid (Fig. 6). The thermal conductivity of a solid is obtained by adding the lattice and electronic components. The relatively high thermal conductivities of pure metals are primarily due to the electronic component. The lattice component of thermal conductivity strongly depends on

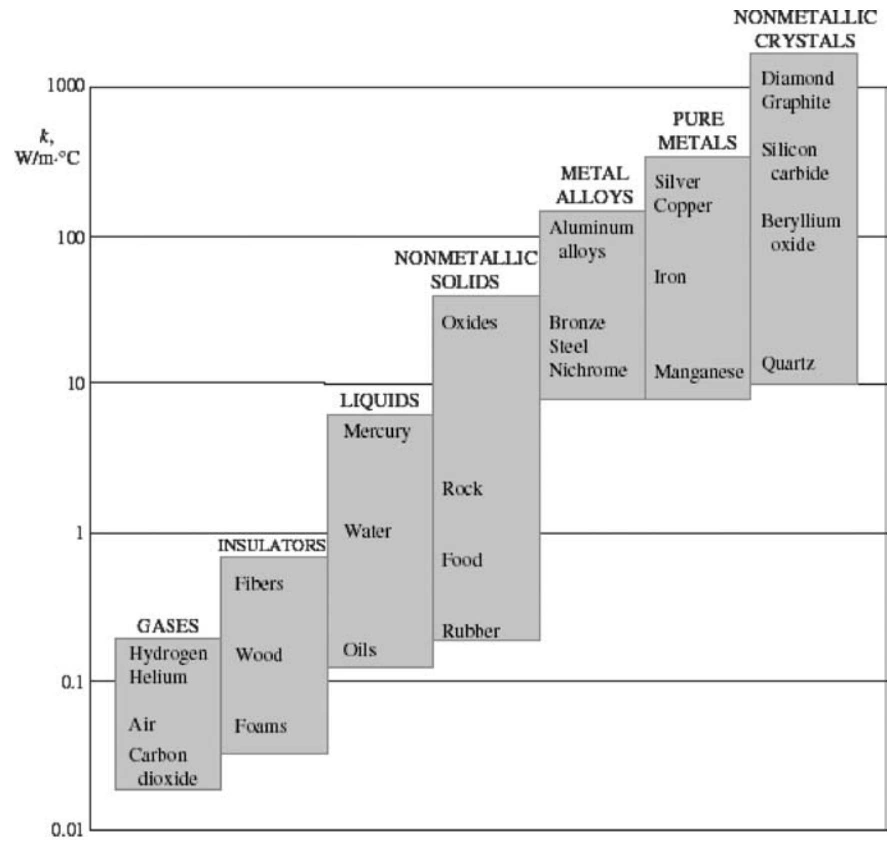


Fig. 5 The range of thermal conductivity of various materials at room temperature.

the way the molecules are arranged. For example, diamond, which is a highly ordered crystalline solid, has the highest known thermal conductivity at room temperature.

Unlike metals, which are good electrical and heat conductors, crystalline solids such as diamond and semiconductors such as silicon are good heat conductors but poor electrical conductors. As a result, such materials find widespread use in the electronics industry. Despite their higher price, diamond heat sinks are used in the cooling of sensitive electronic components because of the excellent thermal conductivity of diamond. Silicon oils and gaskets are commonly used in the packaging of electronic components because they provide both good thermal contact and good electrical insulation.

Pure metals have high thermal conductivities, and one would think that metal alloys should also have high conductivities. One would expect an alloy made of two metals of thermal conductivities k_1 and k_2 to have a conductivity k between k_1 and k_2, but this turns out not to be the case. The thermal conductivity of an alloy of two metals is usually much lower than that of either metal. Even small amounts in a pure metal of "foreign" molecules that are good conductors themselves seriously disrupt the flow of heat in that metal. For example, the thermal conductivity of steel containing just 1 percent of chrome is 62 W/m K while the thermal conductivities of iron and chromium are 83 and 95 W/m K, respectively.

The thermal conductivities of materials vary with temperature. The variation of thermal conductivity over

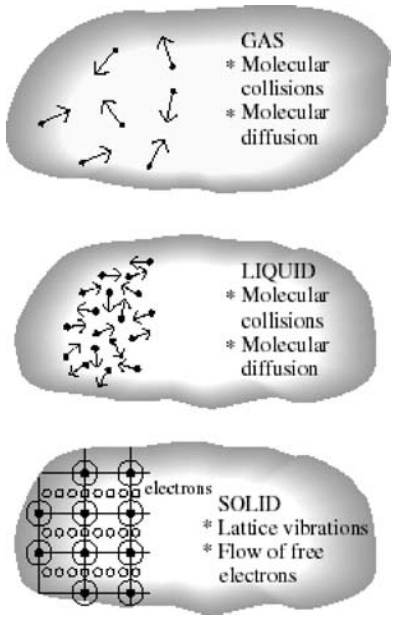


Fig. 6 Mechanisms of heat conduction in different phases.

certain temperature ranges is negligible for some materials, but significant for others. The thermal conductivities of certain solids exhibit dramatic increases at temperatures near absolute zero, when these solids become superconductors. For example, the conductivity of copper reaches a maximum value of about 20,000 W/m K at 20 K, which is about 50 times the conductivity at room temperature. The temperature dependence of thermal conductivity causes considerable complexity in conduction analysis. Therefore, it is common practice to evaluate the thermal conductivity k at the average temperature and treat it as a constant in calculations.

In heat transfer analysis, a material is normally assumed to be isotropic—that is, to have uniform properties in all directions. This assumption is realistic for most materials, except those that exhibit different structural characteristics in different directions (such as laminated composite materials and wood). The thermal conductivity of wood across the grain, for example, is different than that parallel to the grain.

CONVECTION

Convection is the mode of energy transfer between a solid surface and the adjacent liquid or gas that is in motion, and it involves the combined effects of conduction and fluid motion. The faster the fluid motion, the greater the convection heat transfer is. In the absence of any bulk fluid motion, heat transfer between a solid surface and the adjacent fluid is by pure conduction. The presence of bulk motion of the fluid enhances the heat transfer between the solid surface and the fluid, but it also complicates the determination of heat transfer rates.

Consider the cooling of a hot block by blowing cool air over its top surface (Fig. 7). Energy is first transferred to the air layer adjacent to the block by conduction. This energy is then carried away from the surface by convection—that is, by the combined effects of conduction within the air that is due to random motion of air molecules and the bulk or macroscopic motion of the air that removes the heated air near the surface and replaces it by the cooler air.

Convection is called forced convection if the fluid is forced to flow over the surface by external means such as a fan, pump, or the wind. In contrast, convection is called natural (or free) convection if the fluid motion is caused by buoyancy forces that are induced by density differences due to the variation of temperature in the fluid (Fig. 8). For example, in the absence of a fan, heat transfer from the surface of the hot block in Fig. 7 will be by natural convection because any motion in the air in this case will be due to the rise of the warmer (and thus lighter) air near the surface and the fall of the cooler (and thus heavier) air to fill its place. Heat transfer between the block and the surrounding air will be by conduction if the temperature difference between the air and the block is not large enough to overcome the resistance of air to movement and thus to initiate natural convection currents.

Heat transfer processes that involve the change of phase of a fluid are also considered to be convection because of the fluid motion induced during the process, such as the rise of the vapor bubbles during boiling or the fall of the liquid droplets during condensation.

Despite the complexity of convection, the rate of convection heat transfer is observed to be proportional to the temperature difference, and is conveniently expressed by Newton's law of cooling as

$$\dot{Q}_{\text{conv}} = hA_s(T_s - T_\infty) \tag{4}$$

where h is the convection heat transfer coefficient in $\text{W/m}^2\text{ K}$ or $\text{Btu/h ft}^2\text{ R}$, A_s is the surface area through which convection heat transfer takes place, T_s is the surface temperature, and T_∞ is the temperature of the fluid sufficiently far from the surface. Note that at the surface, the fluid temperature equals the surface temperature of the solid.

The convection heat transfer coefficient h is not a property of the fluid. It is an experimentally determined parameter whose value depends on all the variables influencing convection such as the surface geometry, the nature of fluid motion, the properties of the fluid, and the

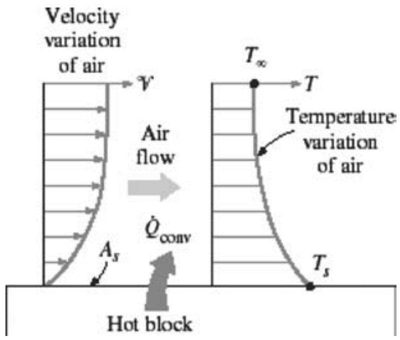


Fig. 7 Heat transfer from a hot surface to air by convection.

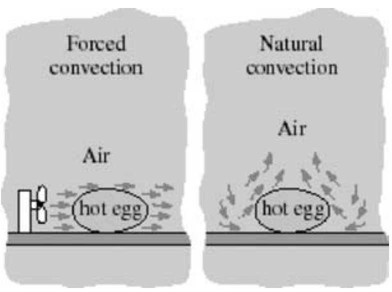


Fig. 8 The cooling of a boiled egg by forced and natural convection.

Heat—Inde

Table 2 Typical values of convection heat transfer coefficient

Type of convection	h, W/m² K[a]
Free convection of gases	2–25
Free convection of liquids	10–1000
Forced convection of gases	25–250
Forced convection of liquids	50–20,000
Boiling and condensation	2500–100,000

[a]Multiply by 0.176 to convert to Btu/h ft^2 R.

bulk fluid velocity. Typical values of h are given in Table 2.

RADIATION

Radiation is the energy emitted by matter in the form of electromagnetic waves (or photons) as a result of the changes in the electronic configurations of the atoms or molecules. Unlike conduction and convection, the transfer of energy by radiation does not require the presence of an intervening medium. In fact, energy transfer by radiation is fastest (at the speed of light) and it suffers no attenuation in a vacuum. This is how the energy of the sun reaches the earth.

In heat transfer studies, we are interested in thermal radiation, which is the form of radiation emitted by bodies because of their temperature. It differs from other forms of electromagnetic radiation such as x-rays, gamma rays, microwaves, radio waves, and television waves that are not related to temperature. All bodies at a temperature above absolute zero emit thermal radiation.

Radiation is a volumetric phenomenon, and all solids, liquids, and gases emit, absorb, or transmit radiation to varying degrees. However, radiation is usually considered to be a surface phenomenon for solids that are opaque to thermal radiation such as metals, wood, and rocks because the radiation emitted by the interior regions of such material can never reach the surface and the radiation incident on such bodies is usually absorbed within a few microns from the surface.

The maximum rate of radiation that can be emitted from a surface at an absolute temperature T_s (in K or R) is given by the Stefan-Boltzmann law as

$$\dot{Q}_{\text{emit,max}} = \sigma_s A_s T_s^4 \quad (5)$$

where $\sigma_s = 5.67 \times 10^{-8}$ W/m^2 K^4 or 0.1714×10^{-8} Btu/h ft^2 R^4 is the Stefan-Boltzmann constant. The idealized surface that emits radiation at this maximum rate is called a blackbody, and the radiation emitted by a blackbody is called blackbody radiation. The radiation emitted by all real surfaces is less than the radiation emitted by a blackbody at the same temperature, and is expressed as

$$\dot{Q}_{\text{emit}} = \varepsilon \sigma_s A_s T_s^4 \quad (6)$$

where ε is the emissivity of the surface. The property emissivity, whose value is in the range $0 \le \varepsilon \le 1$, is a measure of how closely a surface approximates a blackbody for which $\varepsilon = 1$. The emissivities of some surfaces are given in Table 3.

Another important radiation property of a surface is its absorptivity α, which is the fraction of the radiation energy incident on a surface that is absorbed by the surface. Like emissivity, its value is in the range $0 \le \alpha \le 1$. A blackbody absorbs the entire radiation incident on it. That is, a blackbody is a perfect absorber ($\alpha = 1$) as it is a perfect emitter.

In general, both ε and α of a surface depend on the temperature and the wavelength of the radiation. Kirchhoff's law of radiation states that the emissivity and the absorptivity of a surface at a given temperature and wavelength are equal. In many practical applications, the surface temperature and the temperature of the source of incident radiation are of the same order of magnitude and the average absorptivity of a surface is taken to be equal to its average emissivity. The rate at which a surface absorbs radiation is determined from (Fig. 9)

$$\dot{Q}_{\text{absorbed}} = \alpha \dot{Q}_{\text{incident}} \quad (7)$$

where $\dot{Q}_{\text{incident}}$ is the rate at which radiation is incident on the surface and ε and α is the absorptivity of the surface. For opaque (nontransparent) surfaces, the portion of

Table 3 Emissivities of some materials at 300 K

Material	Emissivity, ε
Aluminum foil	0.07
Anodized aluminum	0.82
Polished copper	0.03
Polished gold	0.03
Polished silver	0.02
Polished stainless steel	0.17
Black paint	0.98
White paint	0.90
White paper	0.92–0.97
Asphalt pavement	0.85–0.93
Red brick	0.93–0.96
Human skin	0.95
Wood	0.82–0.92
Soil	0.93–0.96
Water	0.96
Vegetation	0.92–0.96

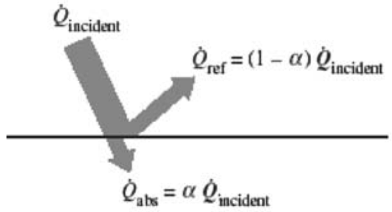


Fig. 9 The absorption of radiation incident on an opaque surface of absorptivity α.

incident radiation not absorbed by the surface is reflected back.

The difference between the rates of radiation emitted by the surface and the radiation absorbed is the net radiation heat transfer. If the rate of radiation absorption is greater than the rate of radiation emission, the surface is said to be gaining energy by radiation. Otherwise, the surface is said to be losing energy by radiation. In general, the determination of the net rate of heat transfer by radiation between two surfaces is a complicated matter because it depends on the properties of the surfaces, their orientation relative to each other, and the interaction of the medium between the surfaces with radiation.

When a surface of emissivity ε and surface area A_s at an absolute temperature T_s is completely enclosed by a much larger (or black) surface at absolute temperature T_{surr} separated by a gas (such as air) that does not intervene with radiation, the net rate of radiation heat transfer between these two surfaces is given by (Fig. 10)

$$\dot{Q}_{rad} = \varepsilon\sigma_s A_s(T_s^4 - T_{surr}^4) \quad (8)$$

In this special case, the emissivity and the surface area of the surrounding surface do not have any effect on the net radiation heat transfer.

Radiation heat transfer to or from a surface surrounded by a gas, such as air, occurs parallel to conduction (or convection, if there is bulk gas motion) between the surface and the gas. Thus the total heat transfer is determined by adding the contributions of both heat transfer mechanisms. For simplicity and convenience, this is often done by defining a combined heat transfer coefficient $h_{combined}$ that includes the effects of both convection and radiation. Then, the total heat transfer rate to or from a surface by convection and radiation is expressed as

$$\dot{Q}_{total} = h_{combined}A_s(T_s - T_\infty) \quad (9)$$

Note that the combined heat transfer coefficient is essentially a convection heat transfer coefficient modified to include the effects of radiation.

Radiation is usually significant relative to conduction or natural convection, but negligible relative to forced convection. Thus radiation in forced convection applications is usually disregarded, especially when the surfaces involved have low emissivities and low to moderate temperatures.

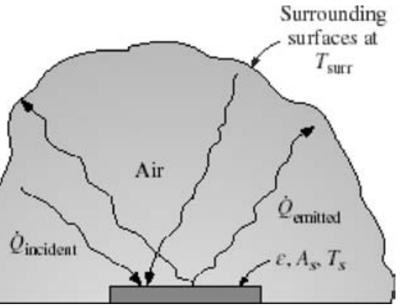


Fig. 10 Radiation heat transfer between a surface and the surfaces surrounding it.

CONCLUSION

There are three mechanisms of heat transfer, but not all three can exist simultaneously in a medium. For example, heat transfer is only by conduction in opaque solids, but by conduction and radiation in semitransparent solids. Thus, a solid may involve conduction and radiation but not convection. However, a solid may involve heat transfer by convection and radiation on its surfaces exposed to a fluid or other surfaces. Heat transfer is by conduction and possibly by radiation in a still fluid (no bulk fluid motion) and by convection and radiation in a flowing fluid. In the absence of radiation, heat transfer through a fluid is either by conduction or convection, depending on the presence of any bulk fluid motion. Convection can be viewed as combined conduction and fluid motion, and conduction in a fluid can be viewed as a special case of convection in the absence of any fluid motion.

ACKNOWLEDGMENTS

The material in this section is abridged from Çengel, Y. A. *Heat Transfer: A Practical Approach*, 2nd Ed.; McGraw-Hill: New York, 2003. Reproduced with permission of The McGraw-Hill Companies.

REFERENCES

1. Çengel, Y.A. *Heat Transfer: A Practical Approach*, 2nd Ed.; McGraw-Hill: New York, 2003.
2. Holman, J.P. *Heat Transfer*, 9th Ed.; McGraw-Hill: New York, 2002.
3. Incropera, F.P.; DeWitt, D.P. *Introduction to Heat Transfer*, 4th Ed.; Wiley: New York, 2002.
4. Kreith, F.; Bohn, M.S. *Principles of Heat Transfer*, 6th Ed.; Brooks/Cole: Pacific Grove, CA, 2001.

High Intensity Discharge (HID) Electronic Lighting

Ken E. Patterson II
Advanced Energy Innovations—Bringing Innovation to Industry, Menifee, California, U.S.A.

Abstract
Electronic High-Intensity Discharge (E-HID) lighting systems are the "high frequency" wave of the future. Traditional and inefficient magnetic HID systems are increasingly being replaced by a new paradigm in lighting, E-HID. In addition to becoming a popular alternative in a lighting designer's arsenal of practical lighting solutions, E-HID has proven to be technologically feasible, economically justifiable, and extraordinarily versatile. However, the differences in E-HID power supplies (ballasts) that are "driving" the surge in new HID lamps, fixtures, and controls are significant. Just calling an HID ballast "electronic" can be misleading in relation to actual performance. Careful analysis is necessary in selecting the "best" E-HID system in terms of energy efficiency, lumens-per-watt (LPW) output, lamp compatibility, and dimming flexibility, as illustrated in this entry.

INTRODUCTION

After over a decade of promise, high intensity discharge (HID) electronic lighting has now become a practical alternative to traditional magnetic-based technology.[1] In particular, high-frequency electronic ballasts for HID lamps have proven to be technologically feasible and economically justifiable, while generating significant energy savings.[2] Because HID lighting systems consume $>4\%$ of our nation's energy, a viable efficiency solution is unfolding as an increasing number of energy efficient electronic ballasts replace old energy depleting magnetic ballasts, thereby reducing our per capita consumption of energy.

However, in addition to energy savings, electronic ballasts can change the entire operational standard for HID lighting. Using this technology, each of the many problematic issues (lumen depreciation, color shift, starts per day, hot restrike times, and lack of controllability) that have plagued HID in the past can be minimized if not totally eliminated.[3] Thus, the impact of electronic ballasts on HID lamps cannot be overstated as amplified in this study. The trend toward electronically ballasted HID lighting is a reality. Indeed, the growing awareness by government stakeholders that electronic ballast usage for every HID lamp should be mandated (as is now done for fluorescent ballasts) will only accelerate this trend.[4]

BACKGROUND

Electronic high-intensity discharge ballasts (E-HID) were introduced into the market in the early 1990s and have long since proven their worth in the market place.[5] Like any new advancement in electronics technology, electronic HID ballasts went through their growing pains—unit failures were rampant and caused many customers to shy away from this technology.

However, as we passed the mid-2000s, E-HID has established itself as a practical, cost-effective, and reliable alternative to traditional HID lighting systems.

The following discussion begins with a primer on electronic ballasts followed by an analysis of the performance of E-HID ballasts compared to magnetic HID ballasts. This comparison incorporates four different ballasts on the market today—Ballast A, an electronic sinusoidal high-frequency ballast with nonoptimal lamp compatibility; Ballast B, an electronic sinusoidal high-frequency ballast with nonoptimal lamp compatibility; Ballast C, an electronic high-frequency sinusoidal ballast with optimal lamp compatibility and avalanche enhanced ignition; and Ballast D, an electronic low-frequency square-wave ballast with nonoptimal lamp compatibility. The purpose is to prove that high-frequency E-HID is a viable and economical solution.

This discussion is also timely because a growing number of end-users are now evaluating the use of E-HID in their facilities and such information contained herein is not easily accessible or available to them.

Electronic Ballast Primer

Dimming Electronic Metal Halide (MH) and Ceramic MH (CMH) ballasts (E-HID) offer significant performance enhancements to virtually every aspect of HID lighting systems—lumen depreciation, color shift, starts per day, hot restrike times, and controllability. Electronic High-Intensity Discharge offers superior energy efficiency, lamp color conformity, lumen maintenance, and lamp

Keywords: Electronic lighting; High-intensity discharge; E-HID; High frequency; Electronic ballast; Sinusoidal; Dimming; Energy efficient; High lumen output; Avalanche ignition.

life. Quiet operation and control of the inherent stroboscopic nature of magnetic ballast and MH lamps are additional advantages.

As recent as 2005, many lighting "experts" thought high-frequency E-HID could not successfully "drive" the new CMH lamps. However, at least one E-HID ballast manufacturer has received certification to operate CMH lamps, opening the door to even greater innovation in ceramic ballasts and lamps in the future.[6]

Types of Electronic Ballasts

There are two types of electronic ballasts (E-HID) for HID lighting. The Low-Frequency Square Wave E-HID ballasts operate at relatively low frequency (100–200 Hz) and have efficiencies of 90%–93%. Although these ballasts offer improved efficacy and lumen maintenance over magnetic ballasts, the low frequency ballasts cannot be dimmed (operated at reduced power) and they consume more energy than High-Frequency E-HID Ballasts.[7] The High-Frequency Sinusoidal Wave E-HID ballasts operate above 100 kHz, offer improved lumen maintenance, can be dimmed (operated at reduced power levels) as low as 35%, and are 92+% efficient. The high- and low-frequency ballasts are suitable primarily for quartz Pulse Start (PS) MH lamps. Some E-HID ballasts are designed to also operate the new, highly efficient CMH lamps, offering the highest efficacy and CRI of all MH systems.

Lamp Ignition/Lamp Performance

Use of a high-frequency sinusoidal electronic ballast brings increased lamp life, improved color, lumen maintenance, and efficiency to HID lighting due, in part, to the characteristics of high frequency operation. The start-up is gentler on the lamp electrodes as the ballasts utilize an order of magnitude (approximately two-thirds) less energy than that of the conventional ballast ignition process. As a result, the life of the lamp and its color rendition will be greatly improved. The electronic ballast isolates the lamp from line voltage fluctuations, which, in-turn, eliminates these line variations as a source for lamp-life degradation.

Dimming and Controls

Lighting consumes close to 25% of our nations' energy. While purchaser's focus tends to be on the initial costs of a lighting system (costs of luminaire, lamp, ballast, and labor), an estimated 88% of the lighting system operating cost is the cost of the energy. As such, daylight harvesting, dimming/load shedding, and other means of controlling energy consumption are becoming popular and will be required by the utilities/regulators in coming years.[4] Because of their ability to be controlled, high-frequency E-HID ballasts are easily utilized in conjunction with dimmers, occupancy, and ambient light sensors to maximize potential energy savings. Most high-frequency E-HID ballasts are dimmable to 50% of lamp power. When dimmed, the light output ranges from 12 to 40%, depending upon the ballast, lamp, and luminaire. One manufacturer can even dim to 35% of lamp power with the backing of multiple lamp manufacturer warranties.[8]

Voltage

Most electronic ballasts have "auto-voltage" and do not need to be "tapped" to operate a particular voltage. They generally handle the range of 208–277, ±10%, 50–60 Hz, AC/DC. The "auto-voltage" feature simplifies and reduces the cost of installation. Most manufacturers also offer a 120 V version, as well. Because some E-HID ballasts hold the lamp to 0.5% of the wattage spec for each 10% variation in voltage, E-HID ballasts provide vastly improved lamp regulation compared to magnetic types.

Starting Behavior

Depending on the ignition process, E-HID ballasts can achieve full power within 45–300 s and full light output within 2–4 min.[9] The time that it takes to go from full power (or energy consumption) to full light output is wasted energy. Due to its proprietary avalanche enhanced ignition, the "C" sinusoidal E-HID ballast has superior time to full power, light output, and hot restrike faster than the standard PS–MH system or "other E-HID" systems.[10]

Efficiency and Operation

Due to the improved lumen maintenance, a facility utilizing E-HID ballasted PS–MH lamps can provide comparable light levels with up to 40% fewer fixtures than a magnetically-ballasted facility. Not only are the light levels and energy efficacy from a LPW (lumens-per-watt) perspective dramatically improved, but the prospect of future environmental waste is reduced due to fewer initial fixtures as well as less frequent relamping over the life of the facility. E-HID ballasts have a power factor (pf) ranging from 90 to 99% and reduced "power distribution losses" due to a much lower THD (Total Harmonic Distortion). Magnetic THD generally runs between 15 and 30% while electronic distortion can be less than 5%.

Electronic High-Intensity Discharge ballast efficiencies of 92%–96% compare favorably to that of reactor and CWA magnetic ballast efficiencies, which peak below 90% when new and become less efficient over life due to heat generation. Of the magnetic ballasts, the more efficient reactor type exhibits the most lamp variability to line voltage variations—for each 5% of line voltage variation the lamp wattage varies by as much as 10%.

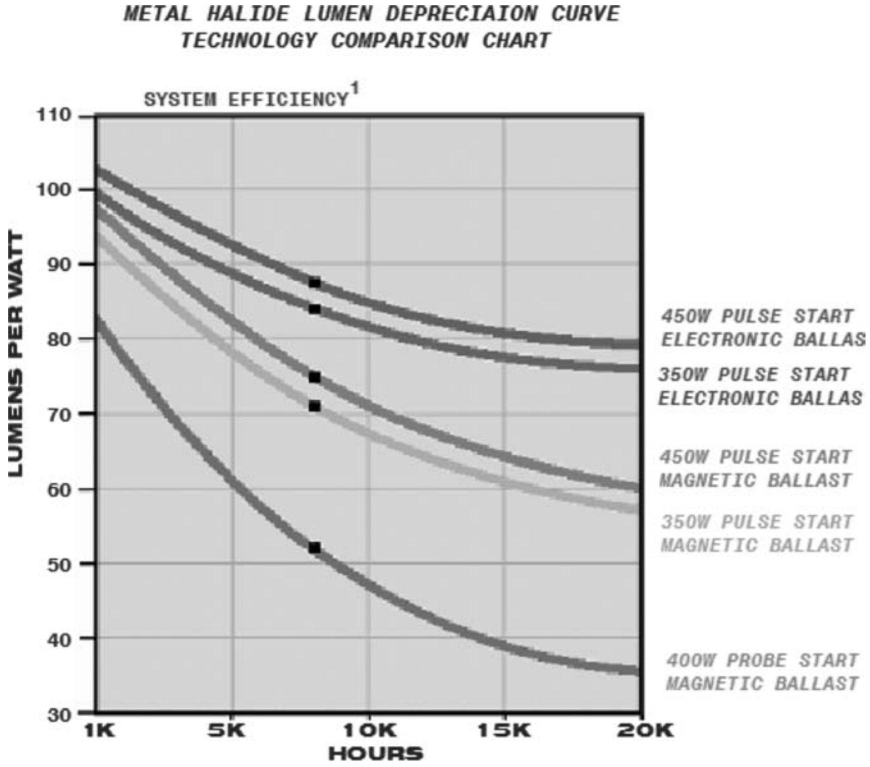


Fig. 1 System efficiency comparison.

As a result, color conformity suffers, lumen depreciation increases, and more regular relamping is required.

Efficacy Comparison

As an example, let's evaluate the number of lamps, lumens, and system watts required to generate 3.4 million mean lumens—enough to light a 60,000 ft^2. warehouse or retail facility at 50 fc. Fig. 1 compares E-HID Electronic Ceramic and Electronic PS systems to Electronic Fluorescent T5/HO, 400 W Magnetic PS and a conventional magnetic 400 W MH M/59 system. The CMH and PS systems light the warehouse using one-third fewer fixtures than the conventional M/59 system. The mean lighting efficacy (LPW) increased 76% with Electronic PS and over 105% with the Electronic CMH ceramic system.

E-HID: LIGHTING THE WAY

Today, many of the major ballast manufacturers are producing some sort of "electronic" HID ballast. However, all electronic ballasts are not alike and most of the manufacturers of electronic HID ballasts fall woefully short on how well their ballast treats the lamp on ignition, operation, dimming, acoustic resonance, and efficacy. In too many cases, promises do not meet expectations as manufacturers have hastened to get in on the flourishing E-HID ballast market. It appears the axiom of "anything is better than nothing" has prevailed.

When one compares costs, the first consideration is "installed costs." Installed costs should include rebates, HVAC savings, demand charges, and system watts at high power (using controls such as occupancy sensors, dimmers, and ambient light sensors), as well as the operating hours at high and low power levels. The incorporation of dimming, system watts (high and low), and percent of time at high and low levels are important drivers to the operating costs of the system.

The utilization of fewer fixtures using a high-efficiency ballast will result in other tangible benefits:

- Reduced HVAC costs due to less system watts and less cooling required; for each 3.4 W saved by the ballast there is a watt saved in cooling/chilling costs.
- Avoided utility company "demand charges" resulting from using controls, demand shedding, and more efficient fixtures.
- Less "power distribution losses" due to a much lower THD and therefore a more energy efficient system; magnetic THD is between 15 and 30%, while electronic is less than 5%.
- Less hazardous waste disposal costs from lamps as a result of less initial fixtures as well as a longer relamping schedule over the life of the facility.

Due to its improved ignition methodology, lumen maintenance, and increased operating efficiencies, high-efficiency electronic ballasts should be included as an option for every new or retrofit magnetic job. At full power, a high-efficiency dimming electronic ballast allows a facility lighting designer to drop down in lamp wattage (between 50 and 200 less lamp watts per fixture than a comparable magnetic system). More importantly, controls allow the system watts and light level to be even lower and fixed until the lamp ages or the luminaire dirt depreciation necessitates increasing power to the lamp. The low-frequency "square wave" ballast cannot offer this advantage because it typically cannot dim the HID lamp.[4]

LIFE EXPECTANCY

The electronic ballast can last as long as a conventional core and coil type (magnetic). At least one type of electronic ballast has a thermal transfer system that avoids the double thermal gradient that is typical in magnetic ballasts.

Most electronic ballast components must first transfer the heat to the air medium then to the housing and again to another air interface (ambient environment). This will lead to an internal thermal gradient of 20°C or higher above the external housing temperature.

However, at least one type of electronic ballast uses a special thermally conductive epoxy to transfer the heat to the housing with a thermal gradient of only 5°C above the housing temperature. It provides many other advantages as well, including vibration and moisture resistance.[6]

IGNITION METHODS

There is a way to start the lamp that is less stressful than ANSI pulse ignition. The ignition methodology of type of one electronic ballast commits a minimum of energy during the beginning of the avalanche process while the work function is low.[8] The lamp is then quickly pushed into thermionic operation in a fraction of the time of the magnetic ballast.

The intermix of the DC offset component and high frequency (3× operating frequency) is optimal to achieve the fastest ignition with the least net energy dump at first break over. Here, timing is everything—push too much, too soon, or too little and it will have the same negative effects on lamp life as a conventional ballast ignition method. Important factors are:

1. Low energy dump at first avalanche
2. Controlled limited crest factor from the start
3. Good dynamic impedance matching

How does this ignition technique of Ballast "C" compare to that of other electronic ballasts? There is a strong similarity in ignition types—certainly they know what works here. However, Ballast "A" and Ballast "B" suffer a delay between first avalanche and thermionic operation. As seen in Fig. 2, the delay in both "A" and "B" holds the lamp in its glow phase much longer than does the methodology inherent in Ballast "C."

With a typical E-HID ballast, the lamp is in glow phase conduction for 1.03 s before the thermionic conduction mode begins. This delay leads to electrode sputtering, blackening of the arc tube envelope, and thus higher lumen depreciation. Lumen depreciation would be exacerbated in

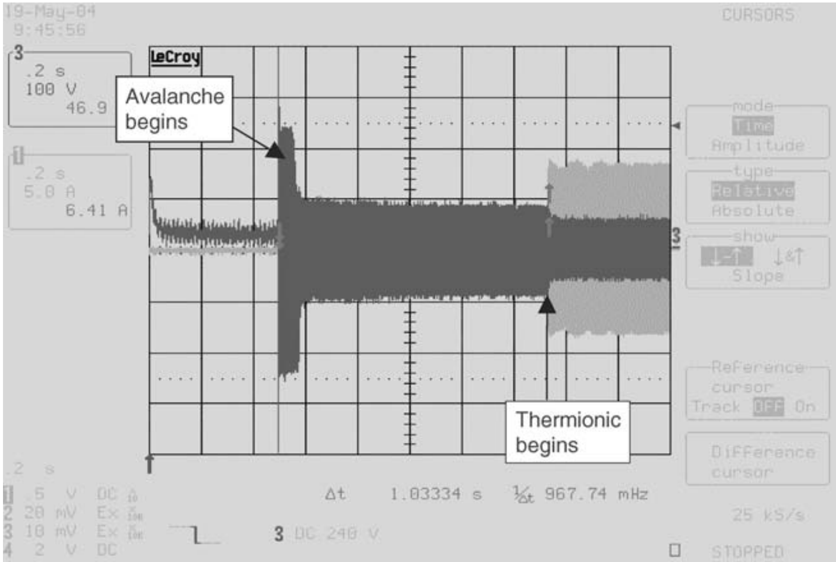


Fig. 2 Typical ballast "A" and "B" ignition @ 0.2 s. per section.

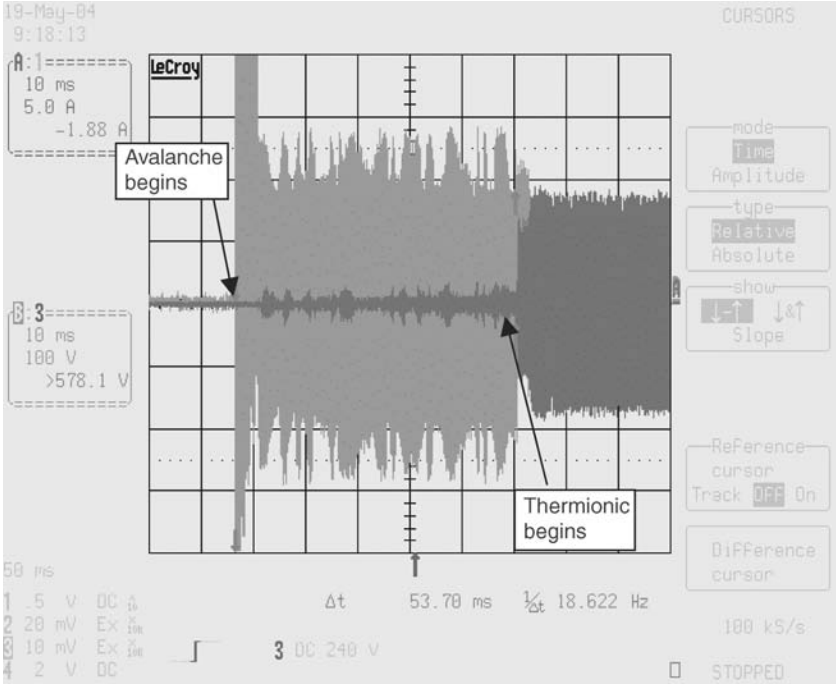


Fig. 3 Ballast "C" ignition @ 0.01 s per section.

applications where frequent on/off cycling is used for energy savings. Furthermore, a slow embedded microcontroller can also adversely affect lumen maintenance.

In Fig. 3, the glow to arc capture for the "C" ballast is 20 times faster than that of the "A" and "B" ballasts. Ballast "C's" glow to arc transition is 0.054 s compared to the "A" and "B" ballasts at 1.03 s each.

FULL-LIGHT OUTPUT

An electronic ballast with an optimized ignition, stabilization, operation, and dimming for HID will provide much better ramp-up/warm-up time while improving the lamp life.

The warm-up time for the "C" ballast is quicker when compared to the "A" and "B" ballasts. This ballast brings the lamp to an equilibrated condition in 30 s. However, the "A" ballast's final operating condition is still not attained at 100 s. It is important to note that this comparison is using the same lamp. Although the calculated current is lower (3.074 amps) for the "C" ballast as compared to Ballast "A" and Ballast "B" (3.192 amps), it comes to full power and light output faster. What's the secret? Better dynamic impedance matching.

MICROCONTROLLERS

Embedded systems have advantages in many applications. However, in HID applications, the lamp phenomena are far too fast and complex for low-cost embedded devices to optimally control. When one compares the time to glow-to-arc transition, the delayed response time of the "A" and "B" ballasts may be attributed to the microcontroller's lag in the acquisition of load data. The "C" ballast is a power supply designed to natively control the nonlinear lamp load and it is not a power supply that has been adapted by the application of external intelligence.

THERMAL CUT-OFF

The idea of a ballast that shuts down when the lamp goes bad sounds like a good idea, but the "C" ballast does not shut down. HID applications are typically critical ones and thus they require a ballast that can run continuously open-circuit without damage. Once a good lamp is placed in the socket, the ballast senses it and begins normal operations. The "A" and "B" ballasts have to shut down in order to save the ballast from damage.

SYSTEM EFFICACY

System efficacy is increased as the lamp is operated in a more ideal thermionic condition. However, this increase assumes that the power delivery does not have time variations in the delivered power to the lamp—something seen in the "A" and "B" ballasts.

When one compares the current and power envelope of the "A" ballast with that of the "C" ballast, the peak of the

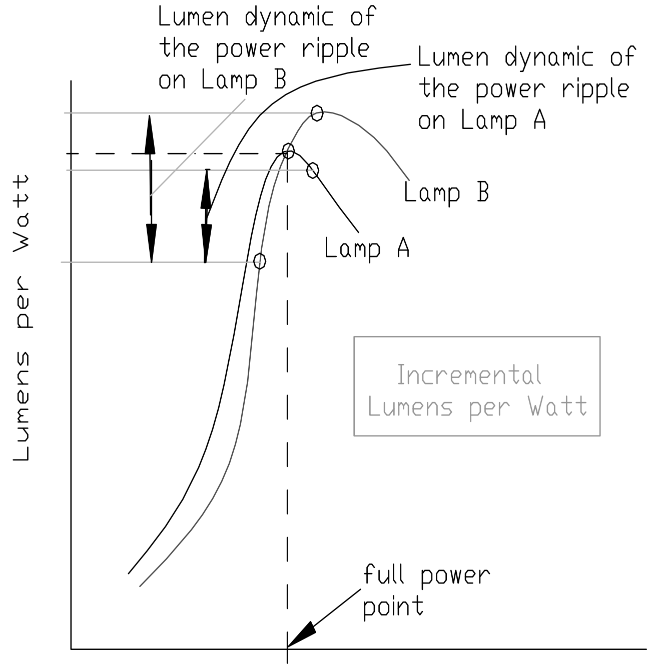


Fig. 4 Power dynamic chart.

power envelope is nearing 1.1 kW while the current peaks are 9+ amps. This methodology of using power pulsing will result in inconsistent performance that will vary depending on the lamp design. New high performance designed lamps, running near their peak LPW conditions, will be affected negatively. This can be better understood by examining the incremental LPW curve in Fig. 4 for two different lamp designs:

This figure shows the results of a power dynamic (change in LPW) analysis of the "C" ballast driving two different lamps. With existing E-HID ballasts (such as Ballasts "A," "B," and "D") the amount of light produced can vary unexpectedly from lamp to lamp. However, with the "C" ballast, there is a total lack of lumen output variation. The black dashed line is the "C" ballast imposed on the lamp curves A and B and yields identical light and LPW. Because the "C"-ballasted lamp has no power dynamic (change of power with time, as described above), a single line fully describes the lamp performance (on both lamps) as well as the expected outcome.

Thus, in the "A," "B," and "D" ballasts, the lumen outcome can vary due to the variation in the change in power over time. When we impose the power dynamic on lamp A, the peak of the RMS envelope exceeds the design point of the lamp. On this graph, the averaged level of the RMS power is still 400 W (on a 400-W lamp), although the peak may sometimes reach 700 W for a short interval. In this lamp, the maximum of the peak of the LPW curve was crossed during the power cycle.

After this peak point, increasing power produces less and less light for each watt increase in power. The net effect on the light output is a lower than expected performance. However, the outcome is very different if the lamp LPW curve peaks at a higher power level than the lamp's rating. Here, the outcome would be to produce more output than expected. The only remaining question is, at what cost? Effects of high power ripple are:

1. Inconsistent and unexpected lamp behavior by different lamp chemistries
2. Different X and Y color coordinates—color conformity dependant on lamp design
3. Different aging effects
4. Inferior lumen maintenance

ARE ALL BALLASTS CREATED EQUAL?

High frequency drives are not all the same at full power operation. There can be many unexpected problems if the ballast behaves in unusual ways.

When one compares one full high-frequency cycle of power delivered by the "A" and "B" ballasts, the energy transfer in the positive half-cycle is quantified at 2.014×10^3 J, while the power transfer in the negative half-cycle is 2.271×10^3 J. The result is a 13% difference in each half-cycle. If this phenomenon is allowed to continue, one electrode will increase in size due to a net transport of electrode material from the other electrode.

As a result of this mismatch in energy transfer from the positive to negative half, lamp performance will suffer

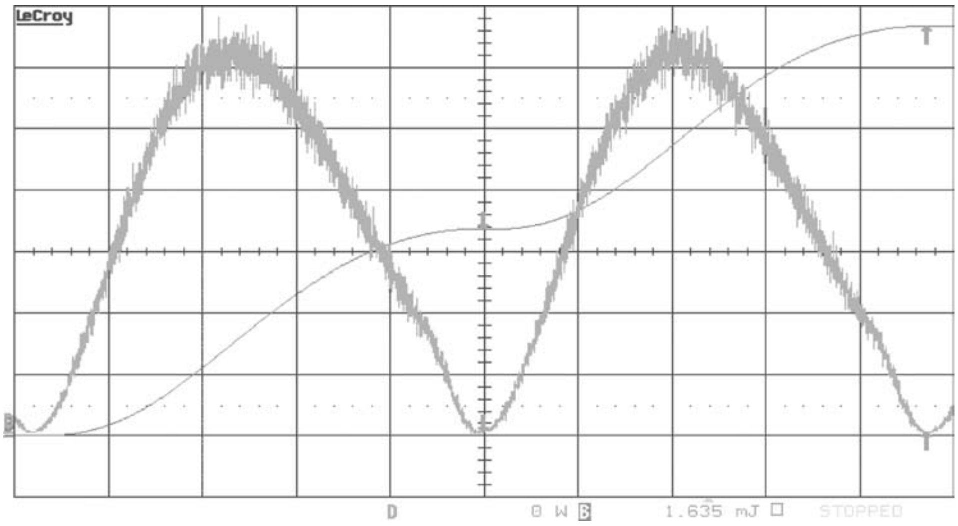


Fig. 5 Power cycle of "C" E-HID ballast.

over the lamp's thousands of hours of operation. An HID lamp operates at its best when there is uniformity and symmetry. Fig. 5 shows a positive and negative half-cycle on the "C" ballast that has mirror symmetry. This symmetry, in combination with the uniformity of the power envelope seen earlier, results in a lamp that will experience an ideal operating condition over its life.

LUMEN DEPRECIATION

As explained earlier, the "C" ballast appears to have been optimized to drive gas plasma loads while others have adapted existing power supply designs to operate lamps. The difference is a holistic system that interfaces with the lamp load during ignition, ramp-up, dimming, and full-power operation. Each one of these elements when precisely choreographed adds or subtracts from the life and performance of the HID lamp.

In Fig. 6, one can see the expected performance of the "C" ballast as compared to ballasts "A" and "B":

At the 8K-hour mark, a real divergence occurs. Ballasts "A" and "B" depreciated 16% at full power while the "C" ballast only depreciated 7.6% at full power. The low-frequency square-wave "D" ballast has an even greater variance than the electronic high-frequency "A" and "B"

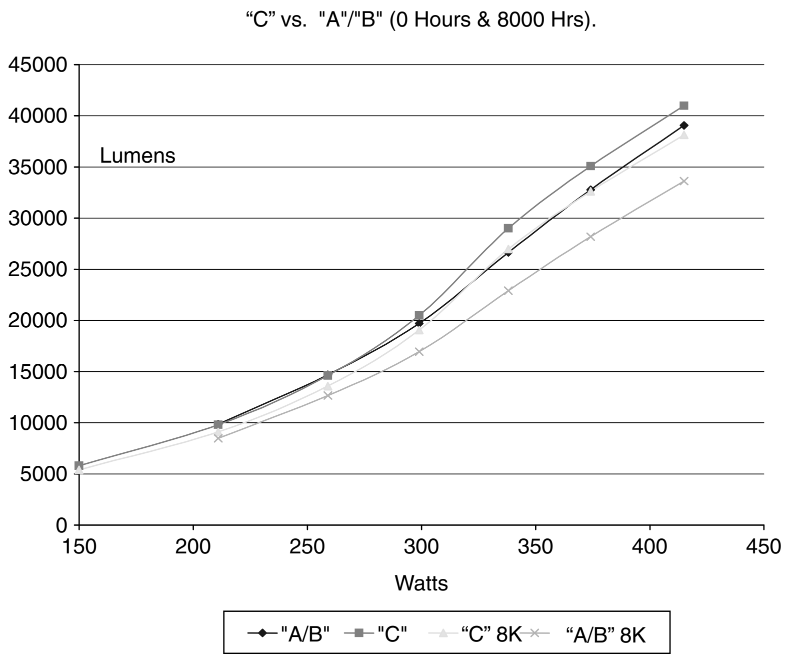


Fig. 6 "C" vs "A"/"B" (0 and 8000 h).

ballasts. The lower depreciation of the "C" ballast is a direct result of optimized operational conditions to reduce or eliminate the mechanisms that destroy the life and performance of HID light sources.

DIMMING LIMITS AND PERFORMANCE

Lamp manufacturers limit dimming to 50% in core-and-coil ballasts and "other electronic" ballasts. For the past decade, the Ballast "C" manufacturer has been working closely with major lamp manufacturers on certification, dimming warranties, and the research and development of new lamps. As a result, this ballast is the first electronic HID ballast to be warranted on both the PS and new ceramic lamp from Osram/Sylvania.

In addition, the "C" ballast is the only one that is warranted to dim to 35% of lamp power by the major HID lamp manufacturers.[6] Multiple lamp manufacturer warranties and an available mix of lamps and lamp types gives the designer options in lighting designs that are unmatched by any other ballast manufacturer. All other E-HID ballast manufacturers have single warranties and are permitted to dim only to 50%.

Also, electronic ballasts do not dim the same or have the same effect on the lamp. There are no standards when it comes to high-frequency dimming. However, there are good principals and practices that must be followed. Dimming has many design aspects that relate to both safety and customer satisfaction:

1. Dimming time should be controlled lest it stimulate the peripheral startle response. This would be dangerous to a machine/forklift operator who would then look to see what changed.
2. Dimming cannot be so slow that the end-user/owner cannot perceive a change. This is important for customer satisfaction and appreciation.
3. Dimming must ensure color uniformity in its transition over the lamp population.
4. Efficiency must be maintained when dimming to achieve and maintain energy savings.
5. Reduced power operation should not significantly degrade lamp color or lamp life.

All of the above design criteria are met in the "C" ballast design. The adaptive power control tracks the lamp throughout its transition through the plasma isotherms.

If the power supply is an adapted type, the dimming progress is achieved through a change-and-wait technique. This occurs when the dimming is artificially controlled for the worst-case lamp that may be experienced, because the dimming power control is not responsive. In other words, if one waits long enough with small transitions, negative resistance effects of the gas curves can be avoided and will never starve the lamp for voltage and extinguish. Ballasts "A" and "B" use this method.

It takes the "C" ballast about 10 s to achieve minimum power and about another five seconds to achieve minimum light. The dimming rate is below that which will stimulate the startle effect and fast enough to be perceived for use of manually operated controls.

Some studies have shown that dimming methodology can actually increase lumen maintenance. One caveat is freezer applications. At −20°C, care must be taken to avoid dimming so low that the freezer temperature has an appreciable effect on the lamp being cooled. This is a situation in which one will not want to dim to minimum power but rather regulate the light output at a predetermined level.

The way ballasts operate the lamp can lead to performance issues. The introduction of dimming electronic HID ballasts allows for a variety of control scenarios to further improve the efficacy of a lighting system and performance of an HID lamp.[11] The utilization of occupancy sensors, daylight harvesters, dimmers, and other energy management tools allows system watts and overall building lighting load to be reduced to as low as 35% of system watts. Table 1 provides the energy consumption (system watts) of the "C" E-HID ballast (full power/dim) as compared to a standard CWA ballasted system.

CONCLUSION

The intent of this entry was not to confuse anyone or to endorse any one manufacturer over another. The first intent was to point out the features, functions, and benefits of electronic HID.[12] The second intent was to show the differences among typical electronic ballasts that are on the market today. These differences are often vast, and just calling an E-HID ballast "electronic" can be misleading in

Table 1 System watts comparison chart between the "C" ballast and a CWA magnetic ballast

Electronic "C" ballast			CWA Ballast
Wattage/lamp: system watts (W)	**Full power (W)**	**Dim (W)**	**CWA (W)**
450	465	173	518
400	413	153	463
350	363	135	403
320	333	125	388
250	263	102	288
200	213	83	230
150	163	65	173

terms of actual performance characteristics as seen in this analysis.

Most consumers seeking a new car shop wisely and compare the features, functions, benefits, gas mileage, etc. before they make a decision. Why, then, should a lighting system end-user buy an E-HID ballast just because the label says "electronic"? As this discussion illustrates, there are significant differences between E-HID ballasts. What is on the inside is vastly more important than what it says on the outside.

REFERENCES

1. Hammer, E. Electronic High Frequency HID Ballasts for eHID Lamps, http://edisonreport.com/HID%20WhitePaper.pdf (accessed on February 20, 2006).
2. *How Do Electronic HID Ballasts Save You Money?*, Office of Energy Efficiency; Natural Resources Canada, Ottawa, Canada, 2006. 1.
3. Beasley, D. High frequency ballasting for HID,. LD + A LIGHTFAIR Preview **1995**, *April*, 27.
4. DiLouie, C. Will electronic HID ballasts take their fair share? The Electrical Distributor **2005**, *August*, 76–77.
5. Houghton, D. Metal Halide—Advances & Improvements. http://gaash.co.il/lighting/html/articles.htm (accessed on January 26, 2006).
6. *Delta's E-HID Ballast now Approved by the Three Major North American Lamp Mfgs.*, Delta Power Supply Press Release, August, 2005.
7. Collins, B. Metal Halide lamp lumen depreciation improvement, US Patent 6,369,522, April 9, 2002.
8. Beasley, D. Essay by invitation. LD + A **2002**, *November*, 8–12.
9. Deurloo, O.; Hollander, J.; Orlov, D., Electronic ballast with ignition and operation control, US Patent 7,019,468, March 28, 2006.
10. Beasley, D. Method for igniting high frequency operated, high intensity discharge lamps, US Patent 5,914,571, June, 22, 1999.
11. Kamoi, T.; Maehara, M.; Mishima, M.; Abe, T. Electronic ballast for a high intensity discharge lamp, US Patent 7,015,655, March 21, 2006.
12. Oomen, M.; Cummings, J.; Diemer, H. Emerging HID Technologies, LD + A **2005**, *March*, 73–75.

HVAC Systems: Humid Climates

David MacPhaul
CH2M HILL, Gainesville, Florida, U.S.A.

Abstract
In humid climates, two of the functions of the heating, ventilation, and air conditioning (HVAC) system are essential—providing proper dehumidification and positive building pressurization—to prevent moisture problems. Although concepts for dehumidification are being discussed and taught to HVAC designers, the HVAC designer must always consider the impact of part-load conditions on the relative humidity levels in the building. The concepts for providing positive building pressurization are less well understood by the HVAC design community. Pressurization must be looked at in each pressure zone at the exterior of the building as well as from a whole-building perspective to ensure the success of the building.

Following a few simple rules when designing HVAC systems for humid climates can help prevent the catastrophic failures seen in such buildings as Florida's Martin and Polk county courthouses in the early 1990s. These buildings suffered rampant mold growth, complaining occupants, and escalating remediation costs because the HVAC system designers did not understand what is required to produce a successful building in a humid climate. Ultimately, the designers and contractors of these buildings were held liable for millions of dollars in remediation costs, which exceeded the original construction costs of the buildings.

INTRODUCTION

Following a few simple rules when designing heating, ventilating, and air conditioning (HVAC) systems for humid climates can help prevent the catastrophic failures seen in such buildings as Florida's Martin and Polk county courthouses in the early 1990s. These buildings suffered rampant mold growth, complaining occupants, and escalating remediation costs because the HVAC system designers did not understand what is required to produce a successful building in a humid climate. Ultimately, the designers and contractors of these buildings were held liable for millions of dollars in remediation costs, which exceeded the original construction costs of the buildings.

Understanding the Functions of the HVAC System

In any climate, the HVAC system designer must consider the four basic functions of the HVAC system:

- Comfort control (temperature, relative humidity, and air motion)
- Ventilation
- Contaminant control
- Pressurization

Keywords: HVAC; Humidity; Contaminant control; Pressurization.

By design, these functions overlap. Ventilation, for example, helps provide contaminant control and pressurization. In humid climates, two of these functions are essential—providing proper dehumidification (a component of comfort control) and positive pressurization—to prevent moisture problems.

DESIGNING FOR PROPER DEHUMIDIFICATION

Dehumidification, or removal of moisture from the supply air, is accomplished in most buildings by cooling the air below its dew point and condensing moisture from the air. In a typical building, the supply air is cooled to about 55°F at the air handling unit. When this air is provided to the occupied space and warmed to a room temperature of 74°F, the resulting relative humidity in the space is 55%–60%, which is comfortable for occupants and well below the 70% relative humidity threshold for mold growth. This process is illustrated in the psychrometric chart in Fig. 1.

Designing a system to handle both the latent (moisture) and sensible loads is taught to, but not necessarily well understood by, most HVAC system designers. However, much has been written recently to increase HVAC engineers' understanding of the problems typically associated with dehumidification in humid climates.[1] For example, discussed the problems that commonly used constant-volume air handling systems have with dehumidification in the Trane Co. *Engineer's Newsletter*.[4] In the *Dehumidification Handbook*,[3] Harriman describes the information HVAC system designers should know to

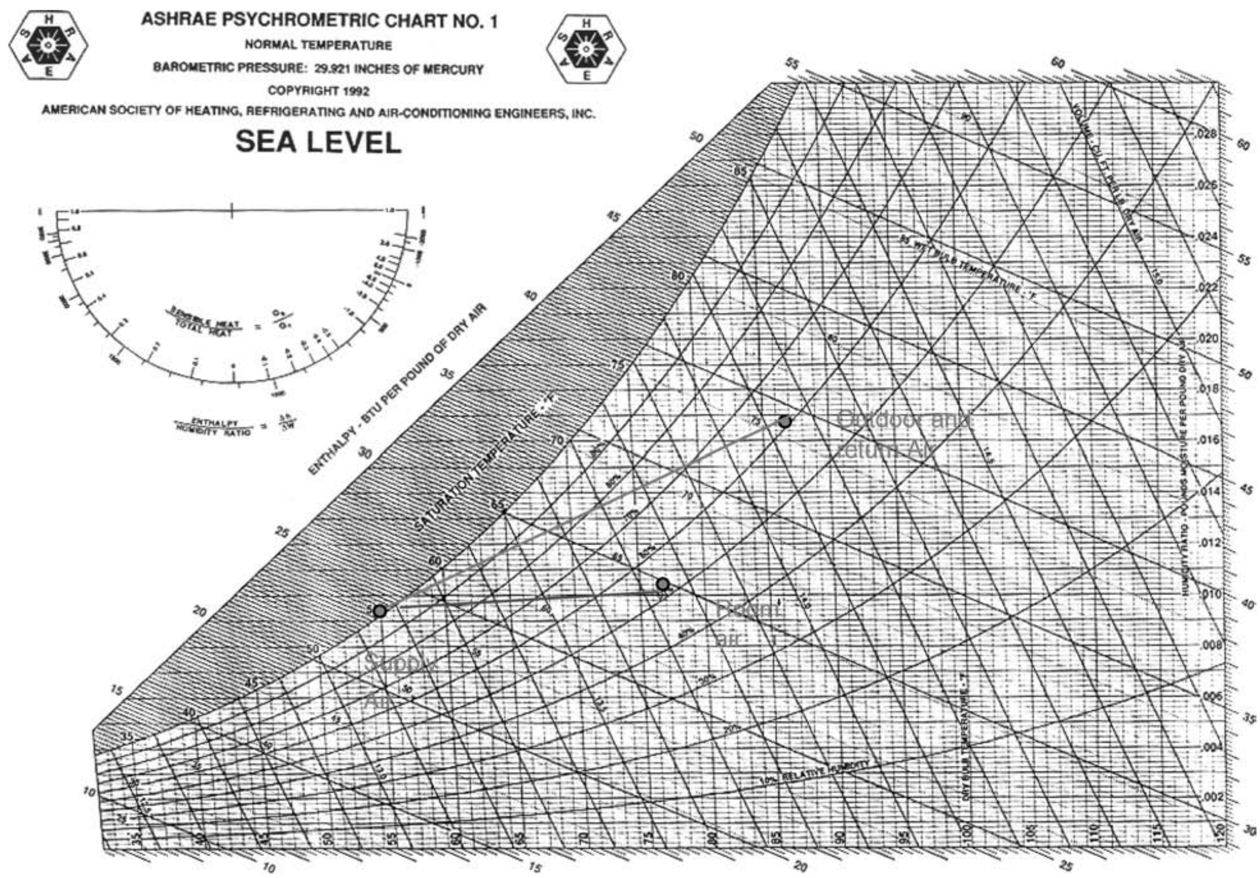


Fig. 1 A typical cooling process shown on the psychrometric chart.
Source: From ASHRAE (see Ref. 1).

prevent dehumidification problems in buildings. Dehumidification problems typically occur because these systems are controlled based on sensible loads, and under part-load conditions, moisture loads may be the largest cooling loads on the system.

CASE STUDY DEMONSTRATING POOR HUMIDITY CONTROL

Consider the vertical stacking fan-coil unit shown in Fig. 2. This type of unit is installed in thousands of hotel rooms throughout the world. In this case, the units were used in a hotel in Honolulu, Hawaii. The hotel owner was required to spend $5 million to remediate a $40 million addition within a year after construction.

As its name implies, a fan-coil unit consists of a supply fan and a cooling coil using chilled water to cool the supply air. As shown in the figure, the unit provides both cooling and ventilation, in this case bringing in 30 ft^3/min of outdoor air.

A simple wall-mounted thermostat controls the valve on the chilled water coil, opening and closing the valve in response to room cooling demands. This type of control responds only to room temperature, not humidity. Suppose that we set the thermostat to 74°F. When the room temperature is 73°F, the thermostat is satisfied even if the relative humidity in the room is 80%—well above the occupant comfort level of 60%. Therefore, the designer must be sure to prevent high relative humidity levels because the system does not actively control humidity; rather, it controls humidity passively as a byproduct of temperature control.

In Fig. 2, the system is providing cooling and removing moisture from the outdoor air. In Fig. 3, the thermostat is satisfied, and the outdoor air provided by the system is brought into the room unconditioned, containing large amounts of moisture. Depending on how long the unit is off, the room's relative humidity may exceed the 60% maximum for occupant comfort.

Fig. 4 shows the expected guest-room cooling loads, both sensible and latent, over the course of a typical day as determined by an off-the-shelf cooling load program typically used by HVAC system designers. Note that the sensible load varies by a factor of two, doubling from the early morning to the late afternoon. The latent cooling load, however, remains nearly constant throughout the day. The system designer must select a fan-coil unit with a cooling capacity at least equal to the worst case load expected. A fan-coil

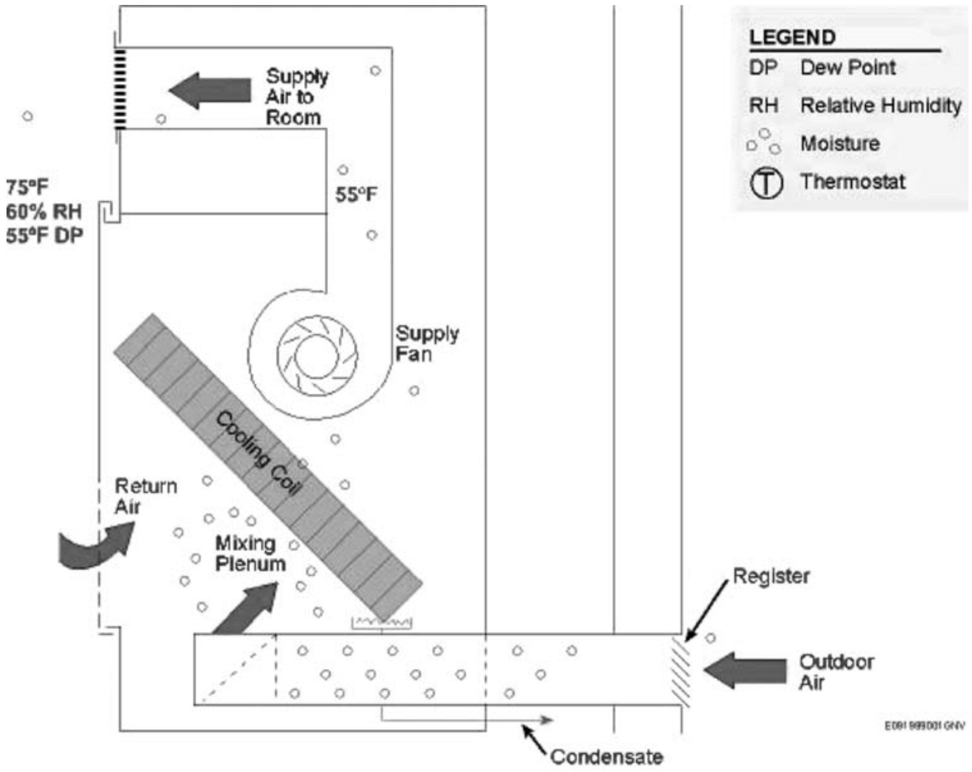


Fig. 2 Fan coil unit used to provide cooling and ventilation in hotel, Honolulu, Hawaii.

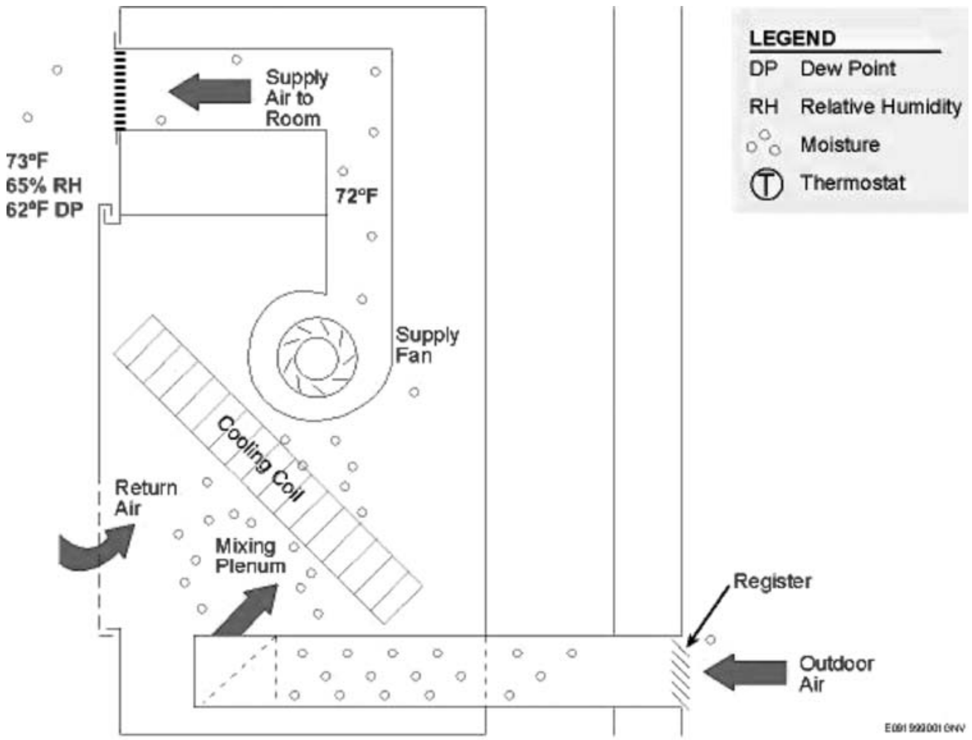


Fig. 3 The fan coil unit does not provide cooling or dehumidification when the thermostat is satisfied.

Heat–Inde

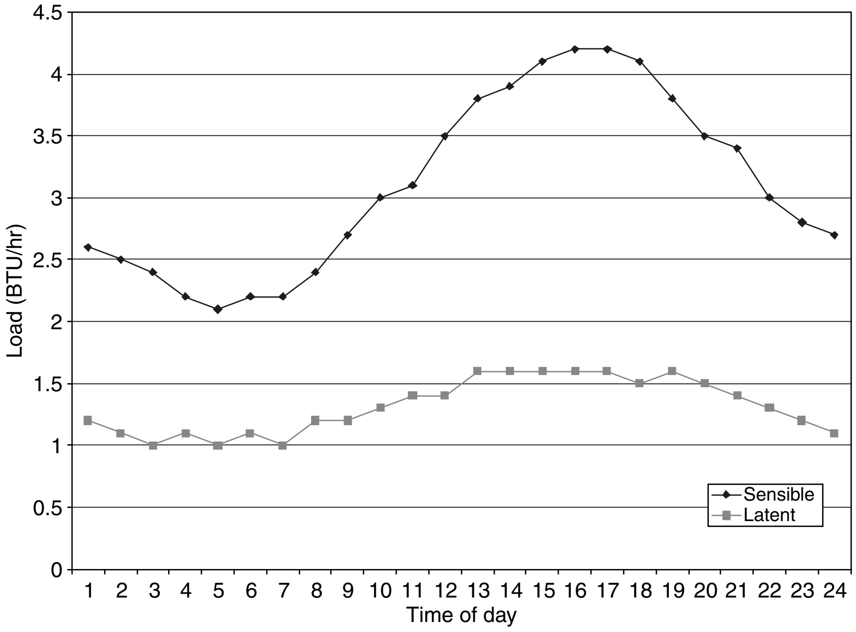


Fig. 4 Hourly cooling load requirements for a typical guest room, Honolulu, Hawaii.

unit with a capacity of approximately 6000 Btu/h is required to cool the room in the late afternoon. Unfortunately, the fan-coil units come in limited sizes, and the designer must select a unit with a higher capacity than needed—in this case, approximately 7500 Btu/h. The unit is oversize by approximately 25%.

To determine the run time (the amount of time the unit provides cooling) of the fan-coil unit for a typical guest room at the hotel, we can compare the sensible cooling required with the sensible cooling capacity of the fan-coil unit on an hourly basis. Fig. 5 shows that the run time ranges from 20 to 40 min/h over the course of a typical day in Honolulu.

In the early-morning hours, room cooling loads are low. The guest room still requires ventilation, and the outdoor air brought in through the fan-coil unit contains a high moisture load. However, the outdoor air is relatively cool—less than room temperature. In Honolulu at night, the outdoor air can actually provide a slight cooling effect even while bringing in large amounts of moisture. Based on the estimated run time of the fan-coil unit, we would expect difficulties controlling the relative humidity level in a guest room of the hotel.

To determine the actual relative humidity levels, data loggers were installed in the guest rooms to record the levels in the rooms over an extended period while guests occupied the room. Fig. 6 illustrates the relative humidity levels found in Room 544. In this room, the relative humidity ranges from 60% to more than 80%! The relative humidity cycles rapidly as the room thermostat opens and

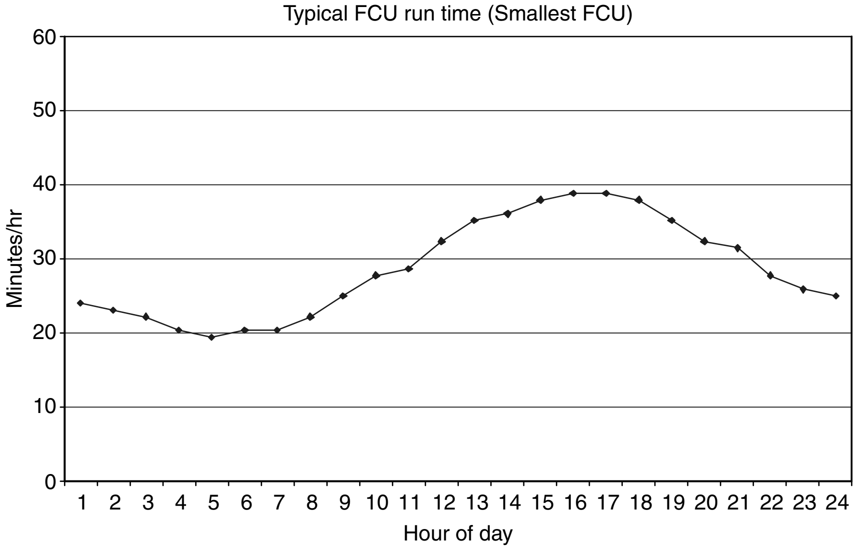


Fig. 5 Estimated fan coil unit run times for a typical guest room, Honolulu, Hawaii.

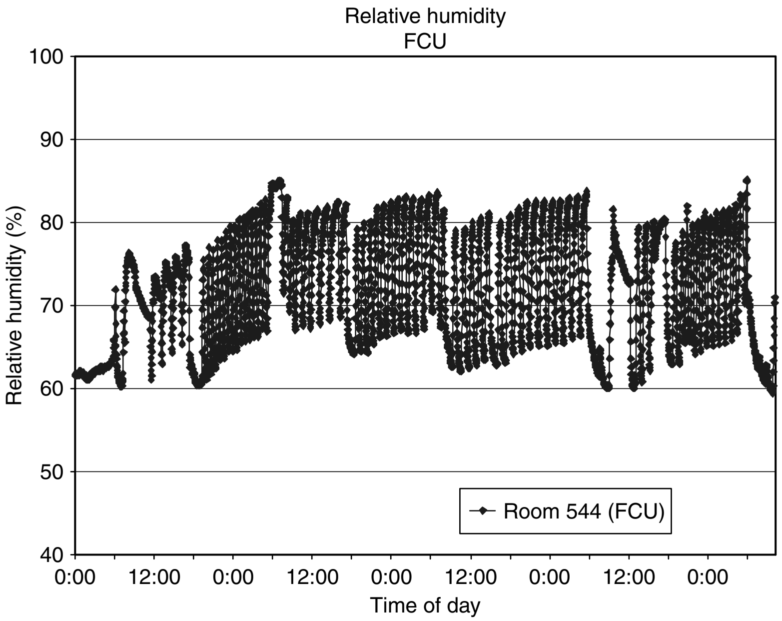


Fig. 6 Actual relative humidity levels in room 544, hotel, Honolulu, Hawaii.

closes the valve on the cooling coil. Needless to say, guests in these rooms found conditions very uncomfortable, and housekeeping staff spent additional time cleaning linens and drapes to remove mold growth caused by the high humidity levels.

In this case, the room's high relative humidity levels were entirely predictable. A simple comparison between the cooling load requirements (both sensible and latent) and the cooling capacity of the fan-coil unit immediately revealed that the unit would not run long enough to remove the required amounts of moisture from the air.

The remedy to this problem was relatively simple. A modified fan-coil unit, as shown in Fig. 7, was installed. In this fan-coil unit, the chilled water coil is separated into two parts. One portion continuously conditions the outside air; the other portion conditions the return air from the

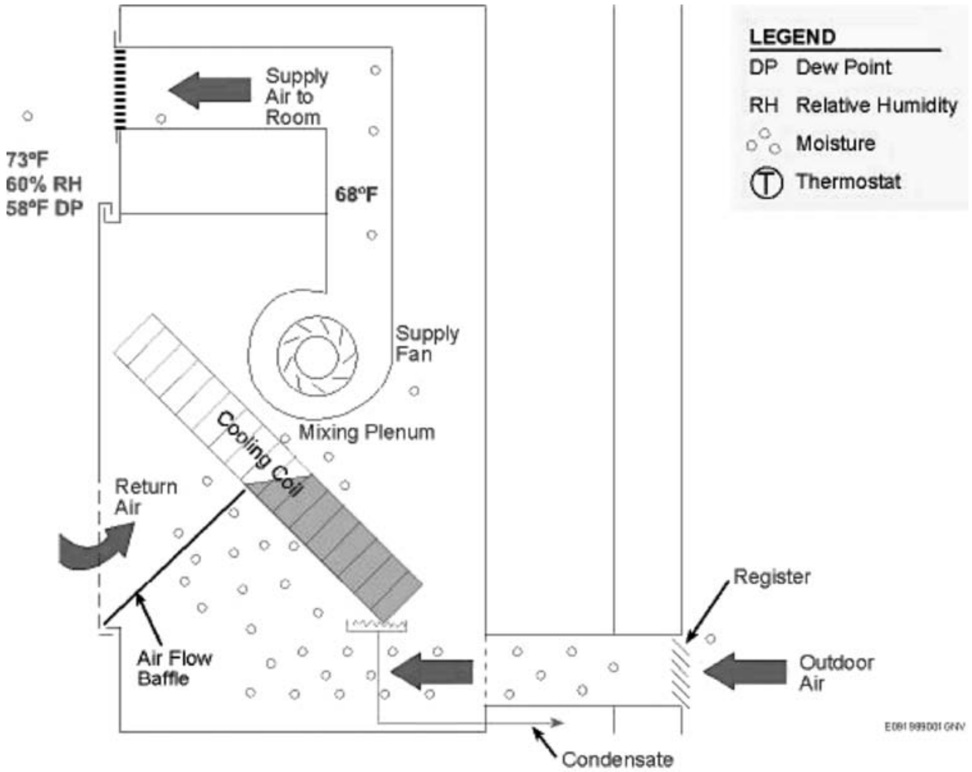


Fig. 7 Modified fan coil unit installed to provide humidity control, hotel, Honolulu, Hawaii.

room and is controlled by the thermostat. The coil itself is not physically separated; the piping providing chilled water is modified to provide the two parts of the coil. Separating the outside air from the return air separates the latent portion of the cooling (the outside air load) from the sensible portion of the cooling (the return air load from the room) and allows for proper humidity control in this small unit, even with the unit oversizing.

This fan-coil unit would have cost about an additional \$100 per room or approximately \$40,000 for the hotel if it had been installed as part of the original construction. Unfortunately, by the time the problem was corrected after construction, extensive remediation of mold growth was required, costing \$5 million.

DESIGNING FOR PROPER PRESSURIZATION

Pressurization in buildings is the second, yet equally important, key issue for producing successful buildings in humid climates. A building must be positively pressurized relative to the outside—that is, some conditioned air from inside the building must exfiltrate through the envelope to the outside. Unfortunately, the required methods for providing proper pressurization are not well understood among the HVAC design community.

Positive pressurization is normally achieved by providing more outdoor, or makeup, air to a building than is exhausted from the building. An HVAC designer typically provides 10%–15% more outdoor air to a building through the air handling units than is exhausted through restroom, kitchen, or laboratory exhaust systems.

Heating, ventilation, and air conditioning designers often treat the building as one large open vessel where air can move freely between zones, believing that the makeup air will go where it is needed in the building. Fig. 8 illustrates the common understanding of pressurization in buildings.

In reality, buildings do not function as open areas at all. Firewalls, floors, shafts, and other full-height walls create a series of pressure zones or vessels within a building. Depending on how the makeup air and exhaust air is distributed throughout the building, most buildings function similarly to the building shown in Fig. 9, containing both positively and negatively pressurized areas relative to the outside, and infiltration and exfiltration are rampant.

The Florida Solar Energy Center (FSEC), in *Uncontrolled Air Flow in Non-Residential Buildings*,[2] coined the term "the smart-air syndrome," describing the belief that air will go where it is needed rather than where it is distributed.

Areas that are positively pressured tend to exfiltrate excess outside air, and areas of negative pressurization have large amounts of infiltration. At best, this infiltration and exfiltration can cause a tremendous energy waste to cool and dehumidify the extra outside air. In the case of infiltration, when moist outside air comes into contact with

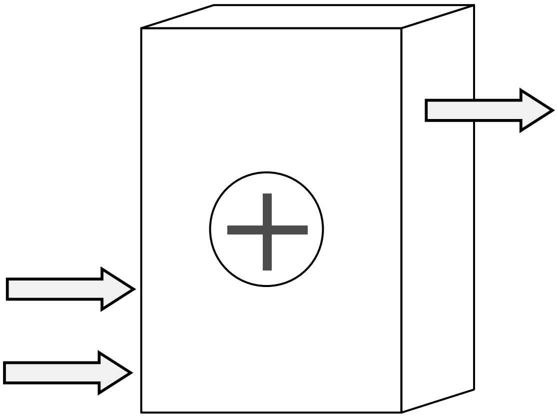


Fig. 8 The common understanding of building pressurization.

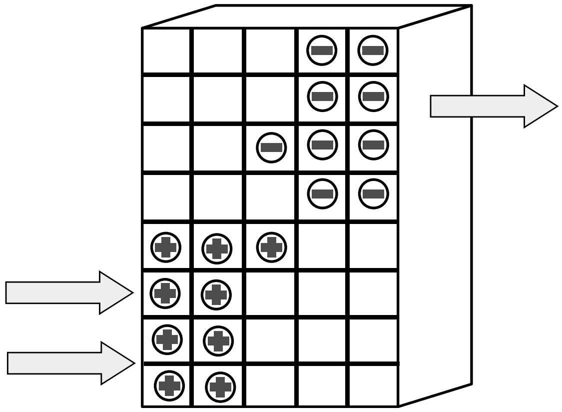


Fig. 9 The actual pressurization in most buildings.

cool building surfaces inside the wall cavity, condensation and moisture accumulation will occur. The amount of moisture accumulation can be quite large. The infiltration of 1 ft^3/min (cfm) of outdoor air in Orlando, Florida, can bring in nearly 20 gal of unwanted moisture per year. In its worst case, infiltration will create severe humidity problems, mold growth, and deterioration of building materials.

When designers refer to the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) *Handbook of Fundamentals*[1] about pressurization in humid climates, they find that it states, "Negative pressures of the indoor space should be avoided." For designers in humid climates, negative pressurization relative to the outside, with its resultant infiltration, must be prevented.

CASE STUDY DEMONSTRATING POOR PRESSURIZATION

The following case study of a hotel in south Florida illustrates what happens when the rules for proper

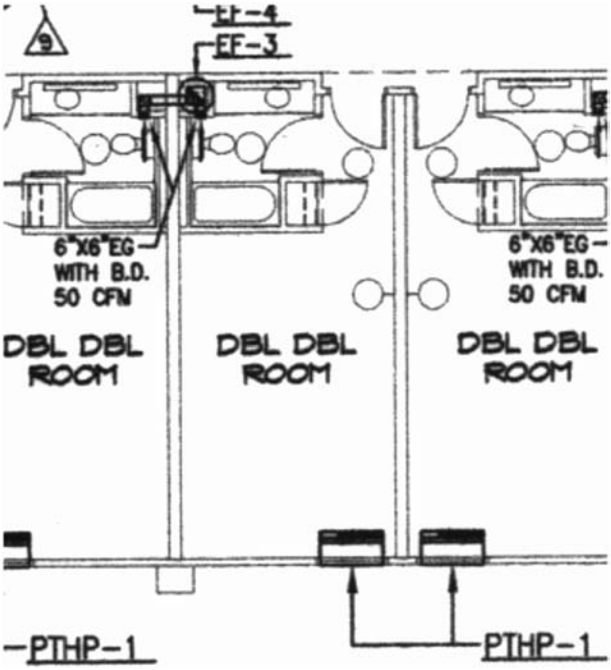


Fig. 10 The layout of the hotel HVAC system.

pressurization are not followed. Hotels are excellent examples because they contain a large number of pressure zones (each room in a hotel is surrounded by fire-rated full-height walls) and generally contain their HVAC systems within the room. In this case, a hotel began to suffer from mold growth, and odors were noted soon after construction was completed.

Fig. 10 shows the design of the hotel HVAC and exhaust systems. The guest room is served by a packaged terminal air conditioner (PTAC) installed in a sleeve through the outside wall. A continuous toilet exhaust fan serving a pair of rooms on each floor is located on the roof. The PTAC unit is scheduled to provide 70 cfm of outdoor air, and 50 cfm is exhausted through the bath fan.

Also, 25 cfm of makeup air per guest room is provided through the air handling system serving the corridor. This makeup air provided to the corridor is intended to serve as a safety factor. Assuming that at least half of the PTAC units would be in operation at any given time, 50 cfm of corridor makeup air would be available to transfer under the guest room door undercut into a guest room with the PTAC unit off, providing the makeup air required for the exhaust system.

To test the designer's assumption, after the hotel began to develop humidity and mold problems, we measured pressures relative to the outside in the guest room and corridor while sequentially shutting down the rooftop-mounted bath exhaust fans. Fig. 11 shows the pressures in the corridor and the guest room. The relative pressure is graphed against the ratio of corridor makeup air to exhaust air.

In this case, the corridor reached a neutral pressurization when the ratio of makeup air to exhaust air is 1.7. The guest room never reached a positive pressurization, even when the ratio of makeup air to exhaust air was nearly 3. Upon further examination, although the test-and-balance report indicated that the exhaust system was properly balanced, significant leakage of air into the exhaust ducts was occurring. Unfortunately, this problem is typical in many buildings. Even with the leakage, the final ratio of corridor makeup air to guest room exhaust air was above 2, and positive pressurization was not achieved as required.

In this case, the owner responded to the humidity problems as soon as they began occurring, and problems were limited through short-term corrective actions.

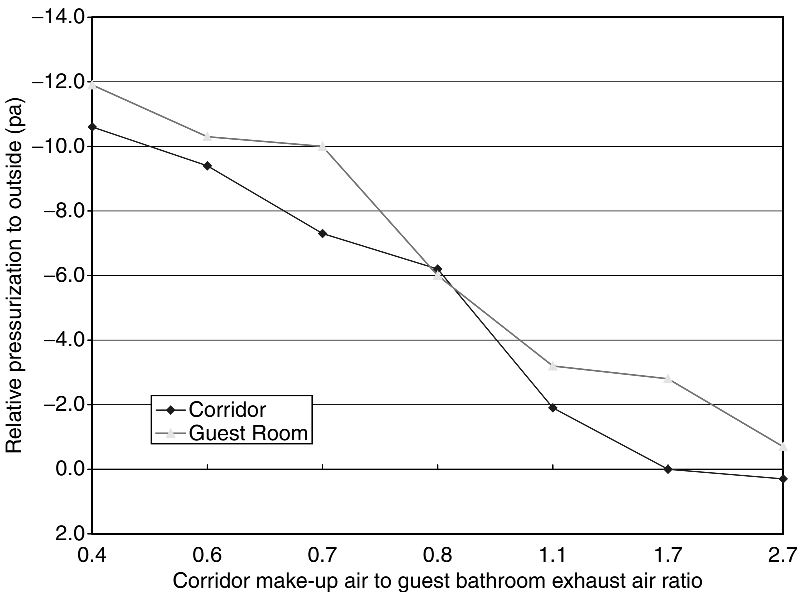


Fig. 11 Relative pressurization of the guest room and corridor.
Source: From CH2M HILL, Feburary 1, 2002.

Fig. 12 Mold growth caused by infiltration in a hotel room.

If the owner had waited to correct these problems, serious mold growth and damage could have occurred. Fig. 12 shows the result of waiting through one summer season before correcting the problems: large amounts of mold growth on the back side of the gypsum wallboard. This piece was removed from the partition wall between rooms approximately 4 ft from the perimeter wall. Extensive and costly remediation work was required at this hotel.

CONCLUSIONS

Providing proper dehumidification and maintaining positive pressurization in all zones of a building is critical to ensuring the success of the building in a humid climate. Although methods for dehumidification are being discussed and taught to HVAC designers, the designer must always consider the impact of part-load conditions on relative humidity levels. For designers, the concepts for providing positive building pressurization are less well understood. Pressurization must be looked at in each pressure zone at the exterior of the building as well as from a whole-building perspective to ensure the success of the building.

REFERENCES

1. American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE). *Handbook of Fundamentals*; ASHRAE: Atlanta, GA, 1997.
2. Florida Solar Energy Center (FSEC). *Uncontrolled Air Flow in Non-Residential Buildings*; Florida Solar Energy Center: Cocoa, FL, 1996.
3. Harriman, L. *Dehumidification Handbook*; Munters Cargocaire: Amesbury, MA, 2000.
4. Stanke, D. Trane Co. Engineer's Newsletter **2000**, *29* (24).

Hybrid-Electric Vehicles: Plug-In Configuration☆

Robert E. Uhrig
University of Tennessee, Knoxville, Tennessee, U.S.A.

Vernon P. Roan
University of Florida, Gainesville, Florida, U.S.A.

Abstract

The introduction of and wide-scale utilization of specifically designed plug-in hybrid vehicles has the potential to reduce the use of petroleum for the production of gasoline fuel by 50%–75%. Further, this can be accomplished without significantly degrading vehicle performance and operability compared to similar conventional automobiles and light truck vehicles (sport utility vehicles (SUVs), pickup trucks, minivans, panel trucks, etc.). The large-scale substitution of electricity produced by pollution-free power plants (nuclear, wind, solar, or renewable biofuels) for gasoline fuel could drastically reduce oil imports, balance-of-payment deficits, and tailpipe emissions. Whether the life-cycle cost of introducing plug-in hybrids on such a massive scale is balanced by the benefits is dependent upon factors beyond the scope of this analysis. However, intuitively, any large-scale reduction in the use of petroleum-based fuels would be expected to improve the energy security as well as the economic and political well-being of the United States. The plug-in hybrid vehicle approach discussed here is one of very few paths forward that appears feasible to begin implementation in the relatively near term with cumulative benefits as the technology penetrates the light vehicle market. The benefits and challenges of using hybrid automobiles to drastically reduce hydrocarbon fuel consumption are discussed below in both qualitative and quantitative terms with the view to show that the proposed approach is feasible, practical, compellingly rational, and thus worth immediate and serious consideration for implementation on a significant scale within a decade.

BACKGROUND

The hybrid vehicle is not a new concept. Hybrids were first conceived and constructed about 100 years ago. Even though they were demonstrated to be a viable concept technically, hybrid vehicles received relatively little attention until the late 1960s when they were considered by some to be a possible way of meeting the newly established emission reduction requirements for automobiles. Early research programs showed that vehicle tailpipe emissions could be dramatically reduced (even for nonplug-in designs) by using hybrid vehicles. However, the real stimulus for pursuing hybrids with an organized effort and significant funding came as a result of the petroleum embargo and resulting energy crisis of 1973–1974. Congress passed Public Law 94-413, the Electric and Hybrid Vehicle Research, Development, and Demonstration Act of 1976, which directed the newly established Department of Energy (DOE) to pursue, among other activities, the technologies associated with electric and hybrid vehicles. Through subsequent DOE funding, many studies were performed and many experimental components, systems, and vehicles were built and tested. However, even as these activities were underway, the energy crisis subsided simultaneously with automobile manufacturers making dramatic improvements in conventional vehicle fuel mileage and tailpipe emission. The introduction and widespread deployment of engine-controlling microprocessors along with continued improvements in exhaust after-treatment led to a near doubling of Environmental Protection Agency (EPA) fuel mileages and more than an order of magnitude reduction in exhaust emissions. As a result, there has been little corporate interest in pursuing the heavier, more fuel-efficient, less polluting, more complex, and more expensive hybrid vehicles. Even so, the considerable work that had been completed showed that hybrids could simultaneously improve fuel efficiency and greatly reduce tailpipe emissions compared to even the dramatically improved conventional vehicles. Consequently, DOE continued supporting some hybrid-oriented research activities as did the not-for-profit Electric Power Research Institute (EPRI) of the electric-utility industry. Electric Power Research Institute is one of the few organizations that has continued an extensive research program for hybrids, with special

☆ The material in this article was originally published as "Using Plug-in Hybrid Vehicles to Drastically Reduce Petroleum-Based Fuel Consumption and Emissions" by Robert E. Uhrig in the Spring 2005 issue of *The Bent of Tau Beta Pi*, quarterly publication of the Engineering Honor Society Volume XCVI, No. 2, Pages 13–19, 2005. That original article has been condensed and modified to reflect changes that have occurred since its publication.

Keywords: Plug-in hybrid; Greenhouse gases; Fuel savings; Home electrical connections; New electrical generation capacity needed; Financial incentives; Transportation model.

emphasis on plug-in hybrid vehicles using batteries that are charged primarily with electricity generated by utilities.

WHAT MAKES HYBRID VEHICLES BOTH RESPONSIVE AND EFFICIENT?

The reason that hybrid vehicles can be designed to accelerate so well is that torque provides acceleration. Torque produced by a gasoline engine increases with engine speed from a low value at low rpm to a maximum (depending on the engine design) perhaps in the 3000–5000 rpm range, after which it falls off somewhat. However, in an electric motor of the type typically used in electric and hybrid vehicles, torque can be quite large, with maximum values occurring at zero rpm and remaining relatively constant during acceleration. Hence, in hybrids, the combination of an electric motor and a gasoline engine together can provide higher torque and better acceleration than would be available in comparable conventional vehicles of equal horsepower. This can occur even though the hybrids usually weigh more and have smaller gasoline engines than the conventional version of similar class vehicles.

The reason all hybrids are so fuel efficient is that the amount of fuel consumed per unit of energy output (specific fuel consumption—pounds or gallons per horsepower-hour) generally decreases with the power level until it reaches a minimum at 65%–85% of maximum power. Thus, a smaller engine running at the optimal percentage of its full power is more efficient and more economical for a given power than a larger, heavier gasoline engine operating at a lower percentage of its maximum power. Some hybrid vehicles also use continuously variable drives that are electronically controlled to operate the gasoline engine at its most efficient conditions over a wide speed range. Furthermore, regenerative braking is used on all types of hybrids. Regenerative braking uses electronic controls to temporarily switch the electric drive motor to an alternator (or generator). The resulting alternator then converts some of the kinetic energy of the decelerating vehicle to electrical energy and stores it in a battery for later use rather than converting it to wasted heat by friction between the brake discs and brake pads.

TYPES OF HYBRID-ELECTRIC VEHICLES

Historically, it was generally recognized that there were two primary power structures for hybrid vehicles—series hybrids and parallel hybrids. However, the evolution of microprocessors and power electronics—allowing virtually unlimited control of the engine and electric drive motor, simultaneously—has tended to blur this distinction. Today, four general types of hybrid-electric vehicles are commonly recognized: Fig. 1, micro hybrids (sometimes called start–stop hybrids); Fig. 2, mild hybrids; Fig. 3, full hybrids; and Fig. 4, plug-in hybrids. These have many common features and similar components, such as regenerative braking, gasoline or diesel engine, electric motor, alternator, battery pack, and central digital-control system. In the diagrams of the four types of hybrids, the size of the schematic representations of the primary components indicates their relative size. As we move from Figs. 1 to 4, the size of the electrical components become larger and their use increases, the gasoline engine becomes somewhat smaller, the performance (acceleration) typically increases, and the onboard fuel economy increases. However, larger electrical components are heavier, more complex, and more expensive. Furthermore, the decrease in the size of the gasoline engine is usually less than the increase in size of the electrical components because the engine must still have sufficient power for acceptable performance after the battery energy is depleted. For these reasons, hybrids can almost always out-accelerate conventional vehicles, especially at relatively low speeds. Drivers generally expect their automobiles to accelerate rapidly yet be economical, and hybrids can provide the desired combination of performance and economy.

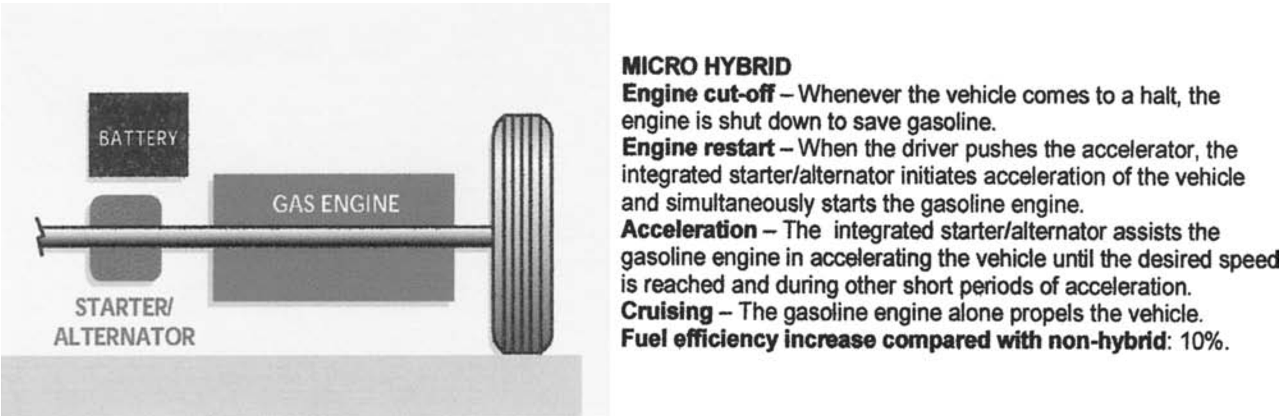


Fig. 1 Micro hybrid.

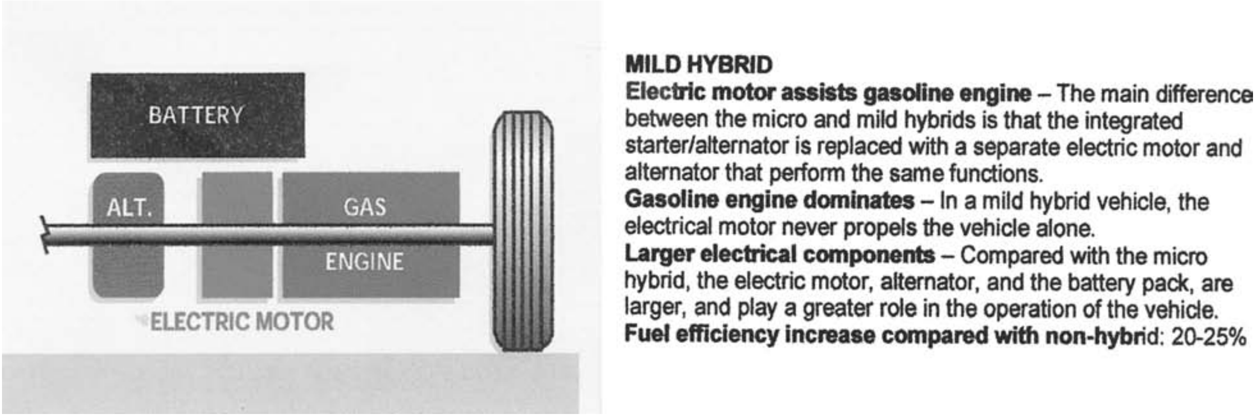


Fig. 2 Mild hybrid.

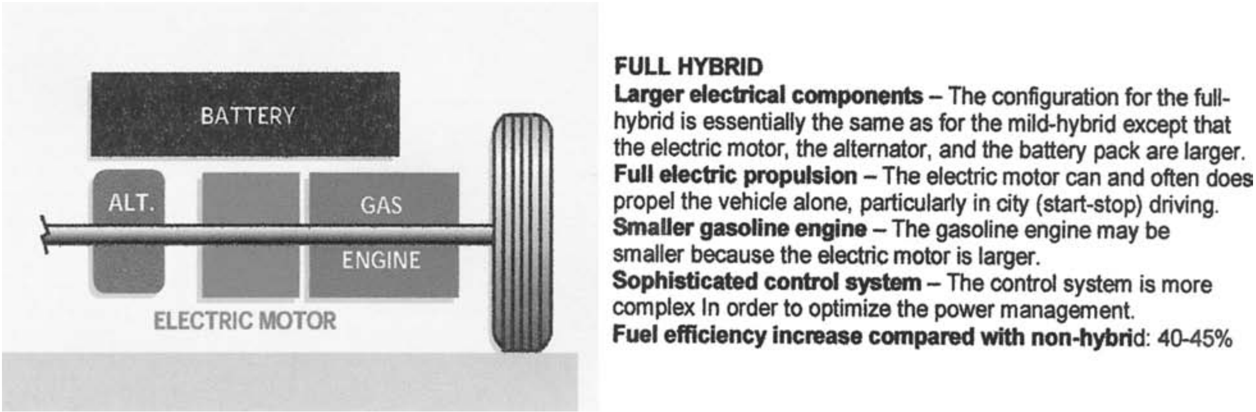


Fig. 3 Fully hybrid.

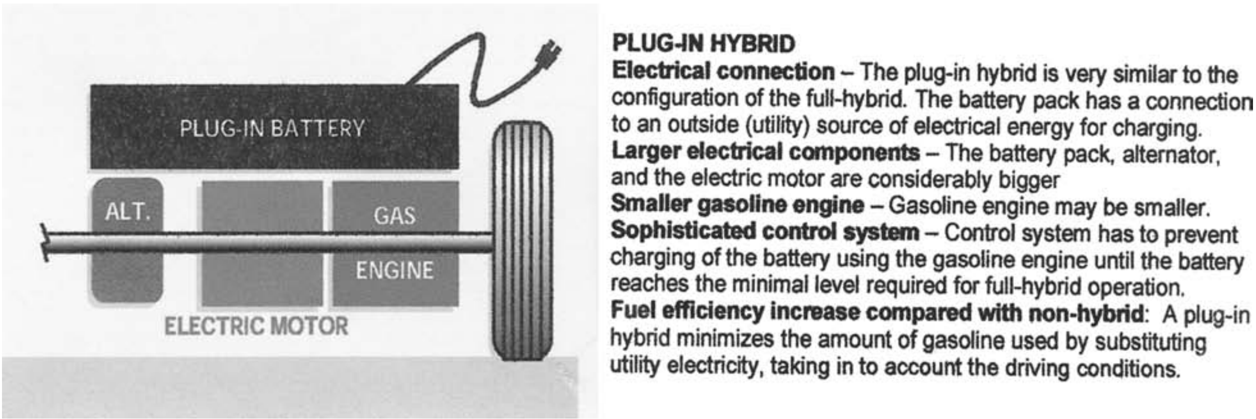


Fig. 4 Plug-in hybrid.

TODAY'S HYBRID VEHICLES

The first modern production hybrid-electric vehicle was introduced to the American public as the Toyota Prius, with the Honda Civic and Insight hybrids following close behind. The success of the Prius, as evident by the long waiting time for delivery, has convinced Toyota to schedule 100,000 Prius models for the U.S. market in 2005 and to introduce two additional hybrid models—the popular light SUV, the Highlander, and a premium SUV, the Lexus RX400h. Toyota and Honda have clearly established that there is a market for hybrid automobiles in the United States. The Prius and the new Ford Escape hybrid are full hybrids. The Civic and Insight models introduced in the late 1990s, the Honda Accord hybrid, and the Dodge RAM Diesel light-truck hybrid scheduled to be introduced in 2005 are mild hybrids. The Chevrolet Silverado and GMC Sierra are micro hybrids that are to be introduced in 2005.[1]

From the standpoints of fuel economy and tailpipe emissions, a somewhat counterproductive trend has appeared with some of the more recent hybrids. In contrast with the Prius that has a 1.5 l, I-4-cylinder 76-hp engine and provides EPA city/highway mileages of 60/51 mpg, the new Honda Accord and Toyota Highlander have 3.0 l, V-6-cylinder 255-hp and 3.3 l, V-6-cylinder, 268-hp gasoline engines, respectively, that provide EPA city/highway mileages of 30/37 and 32/27 mpg, respectively.[5] These large engines, enhanced by the high torque of the electric motors, can accelerate these vehicles from 0 to 60 MPH in 6.7 and 7.3 s, respectively compared to 10.0 s for the Prius (Vehicles with this kind of acceleration are generally considered "muscle" or "hot rod" vehicles, concepts totally inconsistent with the national need for reducing fuel consumption and oil imports). Current advertising for these vehicles emphasize "the high performance with the gas mileage of smaller cars." Assuming the need to reduce use of petroleum-based fuels, a more appropriate theme would be to emphasize "maintaining good performance with much higher fuel mileage." Realistically, however, even large numbers of optimally designed full hybrids will not solve our petroleum import problems.

Full hybrids significantly improve fuel mileage but tend to cost 10%–15% more than other hybrids and somewhat more relative to conventional vehicles. Independent testing by *Consumer Reports* indicates overall fuel mileages of 44 mpg for the Prius, 36 mpg for the comparable-sized Civic, and 51 mpg for the small two-seat Insight, all somewhat less than the EPA-mileage numbers on their window stickers.[1] While these gasoline mileage increases are impressive—typically 40%–45% for full hybrids, 20%–25% for mild hybrids, and 10% for micro hybrids—the differences are not sufficient to have a dramatic impact on the national consumption of petroleum-based fuels. However, a different mode of operation and some redesign of full-hybrids offers a possible means of saving much greater quantities of petroleum. The mode of operation is to utilize battery stored energy until it is nearly depleted without using the engine for recharging. The batteries would receive most or all of their recharging from grid electricity. The redesign involves a larger capacity battery, capable of providing an all-electric range of at least 35 mi. The redesign would also require a larger electric drive motor capable of providing adequate performance without utilizing the engine and an engine capable of providing adequate performance without the electric motor. The control strategy could still use both engine and electric motor for higher performance demands.

One manufacturer, DaimlerChrysler, in cooperation with EPRI, is testing the first of five "plug-in" hybrids (versions of their Dodge Sprinter vans) that are designed for the batteries to be charged by connecting to a grid-based electrical supply.[4] Two gasoline versions will be tested in California and a diesel version will be tested in Kansas.[2] As will be shown below, this approach, if implemented widely, could drastically reduce both fuel consumption and pollution emissions while reducing the cost of fuel for the average automobile owner. Perhaps most important, the widespread deployment of hybrid vehicles would not require changes in driver behavior nor would it require changes in fuel delivery infrastructure. As shown later, large scale use of the plug-in hybrid vehicles would require additional electric power plants and transmission/distribution facilities to deliver additional electrical power throughout the United States.[4]

U.S. LIGHT-VEHICLE TRANSPORTATION STATISTICS

Estimates based on extrapolated DOE–EIA data (unless otherwise indicated, data on U.S. vehicles are from the DOE Energy Information Administration's (EIA) statistics, available on the Internet) from the 1990s indicate that in 2004 there were 225 million light-transportation vehicles in the United States; 133 million were passenger automobiles and 92 million were light trucks (including SUVs, passenger minivans, pickup trucks, and delivery vans). It is further estimated by EIA that on any given day, on average, 50% of U.S. vehicles are driven less than 20 mi. Using these estimates, we can develop a simple model to calculate the potential saving of fuel by the use of hybrids operating in plug-in mode. The model assumes that only the electric motor, operating on batteries charged from electric-utility sources, is used to power a vehicle until the battery has discharged to about one-third of its full stored energy level (assumed to be 35 mi). Beyond that point, the gasoline or diesel engine and electric motor would operate together in the normal full-hybrid mode.

PLUG-IN MODE OF OPERATION

For purposes of this assessment, the standard automobiles and light truck vehicles are grouped and assumed to achieve an overall average of 20 mpg of gasoline or diesel fuel (DOE EIA data shows that in 2003, automobiles averaged 22.3 mpg and light trucks averaged 17.7 mpg; hence, 20 mpg is a reasonable weighted-average value). Let us assume that all of the above vehicles are hybrids capable of the plug-in mode of operation in about three decades i.e., 2035 (this assumption was made in order to evaluate the total potential savings of fuel associated with using hybrid vehicles operating in the plug-in mode. Fuel savings will be reduced in relation to the percent of light vehicles that are not plug-in hybrids). This mode involves charging the batteries of a hybrid overnight, using electricity from an electrical outlet typically in the owner's garage. We assume that when batteries are fully charged, these hybrids can operate using only the electric motor for at least the first 35 mi. For this type of operation, the controls of current full hybrids would need to be modified so as to not use the gasoline engine to recharge the batteries beyond the level necessary to sustain normal hybrid operation. The vehicles envisioned by the authors for plug-in operation are those manufactured by companies to today's standards equipped with normal features such as automatic transmission, air conditioning, and power steering. Hence, a larger electric motor and larger batteries probably would be required, which could possibly lead to the use of a smaller gasoline or diesel engine. Solid-state digital controls capable of optimizing performance and economy while minimizing the use of fuel should make the performance of these vehicles more than competitive with comparable standard vehicles.

MODEL USED TO CALCULATE FUEL SAVED BY PLUG-IN MODE OF OPERATION

The model assumes that each day, one-half of the 225 million light hybrid vehicles operate only for 15 mi on batteries alone while the other half operate on batteries alone for their first 35 mi and then automatically switch to normal full-hybrid mode, in which gasoline powers the vehicles for the remaining miles that day. This means that electrical energy provided to recharged batteries would fuel these vehicles for a grand total of 5.625 billion miles per day (supporting calculations for this and all other derived quantities are available upon request from the author at ruhrig@utk.edu). If the comparable standard (nonhybrid) light vehicles average 20 mpg, then the 225 million light vehicles would use 281 million gallons of fuel to travel 5.625 billion miles per day. Hence, it is theoretically possible (with 100% market penetration), based on this simple model, to replace 281 million gallons (6.7 million barrels) of fuel per day with electricity by using hybrid vehicles operating in the plug-in mode. This represents 74% of the estimated nine million barrels of oil per day now used to produce gasoline for standard automobiles and light truck-based vehicles. Tinkering with assumptions in the model will give slightly different numerical results but will not impact the overall conclusion that it is possible to save a large majority of the petroleum fuel used for light vehicles today through the wide-scale use of plug-in hybrid vehicles.

Recent graphical data attributed to EPRI and Daimler-Chrysler support this quantitative result.[4] Clearly, reductions in both imported oil for transportation fuels and vehicle-emitted atmospheric pollutants would be dramatic with widespread implementation of such full hybrids operating in the plug-in mode. However, realistically, some of the saved fuel would still be needed because in the minimum of three decades needed for full implementation, the number of vehicles and the number of miles driven per vehicle in the United States could increase significantly—perhaps 25%–50%.

FUEL-COST SAVINGS

At a price of $2.00 per gallon, the fuel cost is $0.10 per mile for standard light vehicles averaging 20 mpg. Because a gallon of gasoline contains 36.65 kWh of thermal energy, 1.833 kWh is used per mile. However, the efficiency of an internal combustion engine operating over a range of speeds plus energy losses in the transmission, drive, and tires results in an "overall average gasoline thermal energy to miles traveled efficiency" of about 20%. Hence, the mechanical energy expended at the pavement driving the vehicle is only about 0.37 kWh per mile. If the overall efficiency of the electric drive, including charger, batteries, motor, generator, and drive, is assumed to be 70%, the electrical energy purchased from the utility is about 0.52 kWh per mile. Because the proposed plug-in mode of operation would probably require larger batteries and a larger electric motor, adding several hundred pounds of weight to the vehicle, this value will be increased by 15% to 0.60 kWh per mile. At a price of $0.06 per kWh, the cost of electricity to drive a mile in a hybrid is only $0.036. On average, each vehicle would be driven by electricity 15 miles per day for half the year and 35 miles per day for the other half. Hence, each vehicle would be driven 9125 miles per year on electricity at a electricity cost of $328. Using gasoline at $0.10 per mile would cost $912 per year, so the "fuel" saving for the average hybrid vehicle would be $584 per year.

ELECTRICITY TAX IMPACT ON COST PER MILE

It is inevitable that if electricity becomes a significant source of energy for automotive and light-truck travel it will be taxed by an amount sufficient to recover the tax revenue lost on petroleum-based fuels by governmental authorities at the national, state, and local levels. If we assume that the total tax on these fuels is about $0.35 per gallon and the estimated total consumption of fuel is 103 billion gallons per year (281 million gallons per day), the total tax would be $36 billion per year. Using information provided earlier, the calculated total kWh of electricity consumed in the plug-in mode would be 1238 billion kWh per year. The equivalent tax is about $0.029 per kWh, thereby increasing the cost of electricity used on the road from $0.060 to $0.089 per kWh. Hence, the fuel cost per mile for the light-hybrid vehicles increases from $0.036 to

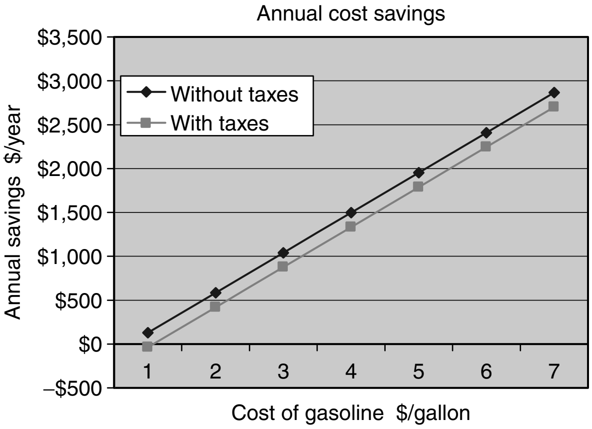


Fig. 5 Annual cost savings using hybrid vehicles.

$0.054 per mile, still little more than half the $0.10 per mile for standard vehicles using $2 per gallon gasoline. Adding this tax would increase the annual cost of electricity from $328 to $493, thereby decreasing the annual savings to $419.

The annual savings for gasoline at $2 per gallon are substantial, but they may not be large enough to justify the additional cost of a plug-in hybrid vehicle. However, if the cost of gasoline increases to $4 or $5 per gallon—prices that are common in Europe today and a realistic possibility in the United States if oil imports are not drastically reduced—the savings become much larger. These annual savings with and without taxes for gasoline prices ranging from $1 to $7 per gallon are shown in Fig. 5.

HOME ELECTRICAL SERVICE REQUIRED

Now let us look at the electrical supply aspects of this system. Light hybrid vehicles that travel 35 mi per day on electricity would use 21.1 kWh per day. If batteries are to be charged in 8 h at night using a 220-V system, the required capacity would be about 12 A. However, charging batteries requires more current initially, when the batteries are deeply discharged, so the peak currents could be double this value. While most modern homes with 200 A electrical service would probably have adequate spare capacity to provide 24 A at 220 V during the night, older homes might require additional electrical capacity. Furthermore, there would be need for plug-in facilities for car owners living in apartments and homes without garages.

ELECTRICAL GENERATION CAPACITY

Even though it is anticipated that battery charging would occur at night when we expect to have excess electrical generating capacity (an expectation that proves to be false), it is important to know the total electrical generating capacity required. Multiplying the 5.625×10^9 mi per day for all the light hybrids using electricity times the 0.603 kWh per mile results in 3.39 billion kWh per day. Charging the batteries in 8 h would require 424 million kWe or 424 GWe. This equals the output of 424 power plants of 1000 MWe size. Because the entire U.S. generating capacity today is about 850 GWe, it is clear that there would not be sufficient spare capacity available at night or any other time to charge the batteries of all the hybrids projected in 2035. Increasing the time for charging to 16 h (including the 3-h span of the three U.S. time zones) would cut the new generating capacity needed to 212 GWe. Hence, some 200 new 1000 MWe nuclear or other nonpolluting plants would have to be built to charge the batteries. New transmission and distribution lines and substations would also be needed to deliver the electrical power. Building 200 new 1000 MWe power plants and associated power delivery facilities in three decades would be a daunting task, but possibly feasible.

UTILITY ISSUES

Clearly, utilities have a key role in the implementation of the proposed plug-in hybrid electric-vehicle transportation system because they must generate and distribute the needed electrical energy. A study by Atomic Energy of Canada Limited[3] of the cost for electricity on the Alberta open market for 2002 showed an average price of $0.0293 per kWh with peaks as high as $0.600 per kWh. Further study showed that the price was less than $0.06 per kWh for 95% of the time and that the average price was $0.0224 per kWh. These prices would probably be lower in the United States, where about 80% of electricity is produced by coal, hydro, and nuclear energy. These competitive costs are low because the capital costs are ignored and the fuel costs for coal, hydro, and nuclear power are quite low. Such a situation is well suited for an interruptible supply mode of operation in which the utility could interrupt the supply of electricity to charge batteries for short time periods (not more than 5% of the time) in exchange for a reduced price for the customer. In most cases, the customer's additional cost, if any, would be the cost of a small amount of fuel. The utility would be relieved of meeting peak demands to recharge batteries when traditional peak utility loads such as air conditioning and heating are high. This means that a utility could delay adding additional generating capacity until its average load increases. It is a win–win situation because hybrid vehicle owners would get reduced electric rates that offset the costs of any extra fuel needed.

INCENTIVES FOR HYBRID VEHICLE OWNERS AND OTHERS

Current Internal Revenue Service (IRS) and some state regulations provide several tax incentives for buyers and operators of hybrid and electric vehicles. Given the large petroleum fuel and pollution reductions of the plug-in hybrids described above, additional incentives such as reduced or no taxes on electricity used to charge batteries at home would seem appropriate—at least in the early years of implementation. The federal government has a large vested interest in promoting any technology that drastically reduces the consumption of transportation fuels, thereby reducing importation of petroleum. Governmental agencies with fleets of light vehicles could be required to increase their percentage of plug-in hybrid electric vehicles on a progressive schedule during a phase-in period, thereby providing a market and incentive for manufacturers to develop and continuously improve plug-in hybrids.

PROBLEMS WITH IMPLEMENTATION

Current full-hybrid designs would require modifications for the proposed plug-in mode of operation for several reasons: (1) current hybrid vehicle electric motors are not powerful enough for satisfactory performance under battery power alone, (2) present hybrid vehicle batteries do not have sufficient capacity to provide the desired power and range for operation with only the electric motor, (3) larger electric motors and batteries will increase the cost and weight of a hybrid vehicle, (4) most batteries are designed for operation in a near fully charged condition and deeper discharge might harm them, (5) battery life (for current batteries) would almost surely be shortened under the proposed mode of operation, and (6) adequate power will be needed for desired accessories of today's light vehicles—air conditioning, automatic transmission, and power steering. These are engineering problems that may be a significant challenge to the manufacturers, but they do not appear insurmountable.

A human-factors issue to be addressed is connecting an electrical source to a vehicle every time it enters a garage or charging station. It is unlikely that a tired worker returning home from a long day at the office will give top priority to reconnecting the power cord to the vehicle. Hence, an automatic docking station that engages an electrical connection would undoubtedly be an desirable feature. Furthermore, an inductive coupling device that avoids mechanical contact would also seem to be a reasonable and appropriate feature of a docking station. Perhaps the biggest impediment to the proposed implementation of plug-in hybrids is time. It takes time to design and introduce even the relatively simple, but needed, changes in the current full hybrids. It takes time for people to accept the changes in operating hybrids compared to their current vehicles. It will a decade or more for a majority of the vehicles in the United States to be replaced. Higher sticker costs are impediments to purchasing new vehicles, even when life-cycle costs may actually be lower. The 30-year period used in this analysis for complete conversion may be unrealistic unless the cost of imported oil remains stable. Temporary reductions of imported oil prices in the past three decades have impeded relatively successful efforts to increase mileage, such as the CAFÉ (corporate average fuel economy) standards for automobiles.

CONCLUSIONS

Given the projected uncertainty in both cost and availability of petroleum, the possibility of replacing up to three-quarters (or even half) of the gasoline and diesel fuel needed for automobiles and light trucks in the United States with electricity by 2035 is extremely compelling. This proposed approach to reducing America's need for petrochemical fuel appears to be significantly simpler and could probably be accomplished sooner and much more inexpensively than any other approach presently under consideration. Indeed, it is probably the only technology that could be implemented quickly enough to have a significant near-term impact on oil imports—e.g., perhaps 10%–20% reduction within a decade relative to business-as-usual.

While the 30 years of the next three decades may seem like a long time, we are reminded that the 1973 Organization of Petroleum Exporting Countries (OPEC) oil embargo and subsequent energy crisis occurred more than 31 years ago and that little has been done since then to resolve the oil supply situation. Indeed, imports have doubled from about one-third to two-thirds of our total needs in this period. Unless the United States moves ahead decisively to reduce our use of petroleum by all practical and reasonably economic means, our importation of petroleum-based energy that is a critical driving force in our economy will be a far greater problem in 2035 than 2005. The use of hybrid vehicles operating in the plug-in mode is a rational and reasonable alternative that could be implemented in less time than almost any other alternative and it should be explored with deliberate speed. Nothing less than the economic well-being and security of the United States are at stake.

REFERENCES

1. Consumer Reports. What's new—2005 cars, fueling the future. Consumer Reports **2004**, *69* (10).
2. DaimlerChrysler. *Mercedes Introduces Sprinter Diesel Electric Hybrid*. Information provided on internet at www.whnet.com/4x4/sprinter.html, click on Hybrid Page, July 31, 2004.
3. Alistair, M.; Duffy, R.B. *Meeting the Near-Term Demand for Hydrogen Using Nuclear Energy in Competitive Power Markets*, Proceedings of GLOBAL Meeting, American Nuclear Society: New Orleans, LA, November 16–20, 2003.
4. Graham, R. Comparing the Benefits and Impacts of Hybrid-Electric Vehicle Options, EPRI Report #1000349, July 4, 2001.
5. Wired. *Rise of the Green Machine*, April 2005; 97–105.

Independent Power Producers

Hemmat Safwat
Project Development Power and Desalination Plants, Consolidated Contractors International Company, Athens, Greece

Abstract
The historical evolution of the independent power producers (IPPs) is reviewed. The stakeholders of a typical IPP under a long-term power purchase structure are described. The roles of the various participants in a project of this type are summarized. The various stages of an IPP—from development to plant construction and plant operation—are described.

INTRODUCTION

An independent power producer (IPP) is a result of privatization and economy liberalization in infrastructure in various parts of the world that started in the early 1980s. Since that time, IPPs have taken various forms as the electricity markets evolved in various countries and regions. The evolutions are closely related to dynamics of structural/regulatory, economic/financial, technological, and environmental conditions. Thus, in planning and implementing an IPP, a multitude of considerations have to be thoroughly attended to.

Historical

Historically in the United States, utilities have operated under a regulated environment. Under this regime, a municipal, cooperative, or privately owned company had a monopoly on the supply of electricity to all customers in a specific region or area. Customers paid tariffs set by regulatory bodies that used review processes that essentially allowed the utilities to earn set returns above their reasonable costs. As a result of deregulation in several industries in the 1980s (e.g., the breakup of AT&T, the deregulation of the airlines, etc.), electricity regulation was criticized, and claims of large inefficiencies in the electricity industry were voiced. Various moves to encourage/introduce incentives for the nontraditional generation started to come about. For instance, in cogeneration plants where a producer in an industrial/agricultural setting could produce electricity, the utility serving the area where the producer built the plant would be obliged to buy the extra electricity that the producer generated under set rules. The trend expanded, and the small cogeneration plants got bigger and spread across the country, accompanied by changes in state laws. These led to the introduction of IPPs that would primarily generate electricity for sale as their main business.

An IPP is a nonutility power-generating entity that typically sells the power it generates to electric utilities or large users at wholesale prices. IPPs are unlikely to be franchised to distribute electricity or to own transmission lines for transmitting their generated power. The IPPs usually sell their power under a power purchase agreement (PPA) with the incumbent utility, the single purchaser, or large customers. It is common to see the PPA take the form of build own operate transfer (BOOT), which permits an investor to "build," "own," and "operate" the generation facilities and then "transfer" them to the host purchaser upon the termination of the contract. In fully restructured markets, IPPs may also operate on a purely commercial basis in which generators can offer their production in the daily wholesale energy markets. In either case, there is a tacit assumption that the IPPs are assured access to the transmission system on a nondiscriminatory basis.

The contribution of IPPs to the energy market can be huge. For example, IPPs owned and operated approximately 40% of U.S.-installed generating capacity in 2003.

It is noteworthy that in the U.S. context, the term IPP excludes qualifying facility (QF) from the definition. QFs are also nonutility power producers that meet certain operating, efficiency, and fuel-use standards set forth by the Federal Energy Regulatory Commission (FERC). They are relatively small power producers that use renewable and alternative energy resources like hydro, wind, solar, geothermal, municipal solid waste, or landfill gas to generate power. QFs, which may also include cogenerators, sell their output to the electric utilities, which must purchase this power at prices approved by the relevant state regulatory bodies. The price that the utilities pay is based on the cost that the utilities avoid by purchasing QF-produced power instead of generating or procuring it from third parties. These arrangements are governed by the Federal Public Utility Regulatory Policies Act (FPURPA) of 1978, which was enacted to stimulate the

Keywords: Independent power producer (IPP); Build own operate (BOO); Build own operate transfer (BOOT); Private power.

use of alternative and renewable energy sources in the generation of electric power.

Outside the United States, the liberalization of the economies of various countries led to a trend of shifting electric utilities from public (government) ownership to a private or mixed (government-private) ownership structure.

In the Middle East, because of the lack of water, the emergence of independent water and power producers (IWPPs) is noted. The IWPPs have grown to megaprojects involving large-capacity "power and seawater desalination."

It is noted that electricity service involves three basic elements: generation, transmission, and distribution. This section is concerned primarily with generation (i.e., the IPP is part of electricity generation). An IPP builds and operates a generation asset, and the electricity generated flows through other transmission/distribution systems owned/operated by others to reach the end consumers: bulk users such as large industries, industrial/commercial users, and households.

IPP PERSPECTIVE

From the perspective of an IPP, its interest starts in recognizing an opportunity to build a power plant or, in some cases, acquire a generation facility. The demand for the electricity to be produced is of paramount importance. This, coupled with the ability of the IPP to produce electricity at a competitive price while earning acceptable returns on its investments, all within a risk posture that is acceptable to those involved in the ownership of the enterprise, provides the impetus for starting a project. The perspective considers the various stages of the project, including the planning, the building of the facility, and the operating of the facility when it reaches commercial operation for the duration of the planned term or life of the facility.

In some instances, the opportunity starts from a competitive bidding process initiated by a government agency in a specific country. In other instances, directly negotiated arrangements are followed.

VARIATIONS IN IPPS

Two distinctive forms of IPPs are found—long-term power purchase (LTPP) and merchant (M) types—as well as mixes of the two types (LTPP+M). In the former, the IPP enters into an exclusive arrangement with an off taker who purchases all the electrical energy that the IPP produces. The two sides enter into a PPA that sets the obligations/rights of the two sides. Agreements in addition to the PPA involve other stakeholders beside the IPP and the off taker. This arrangement is generally followed where a single-buyer model is in effect and this single buyer is the off taker, as designated by law in the jurisdiction.

In the case of a merchant-type IPP—generally found in fully restructured markets—the IPP builds and operates a power plant that is dispatched or that is responding to market demand as it decides in accordance with economic merits of its own choice. In a mixed-type IPP (LTPP+M), the IPP commits to sell part of its production under a long-term arrangement, and with the other part, it exercises its own prerogative to generate or sell as market conditions meet its economic criteria.

THE IPP AND STAKEHOLDERS

An LTPP-type IPP is chosen for the purpose of introducing typical stakeholders associated with a project, referred hereafter as building a power plant and operating it for the duration "term."

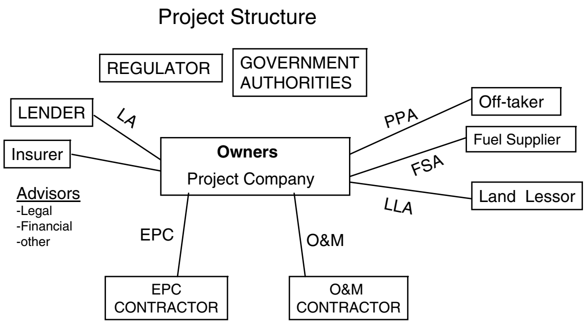
Fig. 1 depicts the stakeholders typically involved in the project. The figure also shows the agreements that are typically entered into for a LTPP.

Project

For the purposes of the discussions introduced hereafter, a project is defined as the execution of all the planned activities, starting from the idea of building a power plant to completing the plant, operating the plant, and selling the electricity generated from the plant.

The sections titled "Project Development", "Plant Construction", and "Plant Operation" provide further insights into the stages of the project.

As noted above, a number of agreements are completed for a project. These are referred to as the project agreements.



Project Agreements
PPA: Power Purchase Agreement, FSA: Fuel Supply Agreement, LLA: Land Lease Agreement
LA: Loan Agreement
EPC: Engineering, Procurement and Construction, O&M: Operation and Maintenance

Fig. 1 Stakeholders and project agreements.

Owner/Developer

The owner is the investor who will own the plant. Typically, the owner/investor invests in the form of equity as a portion of the total investment in the project. The other portion is a debt portion. The owner/investor may be a single entity, but typically the owner is a group of investors with different equity sharing. There also may be principal investors, together with investors with minority shares.

The developer is the manager of the effort to realize the project from its inception to (typically) the financial close, which triggers the start of construction of the facilities for the project.

Quite often, the principal investors act as the developer, though in some projects the developer may be a nonowner who acts independently until a certain stage in the development and then sells the work up to that point to a prospective owner; the developer may or may not continue to manage further the remaining development for the owner.

The owners form a project company that undertakes the project when necessary agreements among the various concerned stakeholders are completed and are in force, as applicable.

The project company is the special-purpose company that is formed by the owners/investors that undertake the project in accordance with the owners' shareholder agreement.

Off Taker

In a typical LTPP, the IPP enters into a long-term PPA in which the IPP is the seller of the electricity it generates in its facility and the off taker is the buyer of the electricity from the project company.

The PPA addresses the obligations and the rights of the two parties. Fig. 2 lists the key topics typically covered under the PPA.

Generally speaking, the owners/investors prepare the PPA and negotiate its terms on behalf of the project company (until it is formed) with the off taker. After the project company is formed, it becomes the responsible party as the seller of the electricity or the IPP.

Fuel Supplier

For a fossil plant, the owners/investors—and, upon financial close, the project company—enter as a buyer from a fuel supplier into a fuel supply agreement (FSA). For a LTPP, the items listed in Fig. 3 typically are addressed in the FSA.

Regulator and Governmental Authorities

Prior to the formation of the project company, the owners/investors and developer undertake all necessary steps to obtain all necessary permits, licenses, and consents from the government authorities in accordance with all applicable laws. They prepare and file the applications, and participate as needed in the presentations and reviews of the government bodies for these applications. When the permits/licenses/consents are granted, the project company acts in accordance with the specifics of the permit/license and takes steps to renew them in accordance with the conditions of the permit/license.

Environmental Agencies

The owners/investors initially and upon formation the project company have to attend to all of the requirements of the applicable laws and the related procedures, as dictated by the environmental agency. These relate to emissions, waste, and noise. The World Bank guidelines have emerged as the standard that is used in most countries, and these guidelines now serve as the generally accepted standard.

Power purchase agreement

- Off-take / Capacity / Phases (if any)
- Term
- Charges:
  - Capacity charges
  - Energy charges (Fixed & Variable)
  - Availability-Penalties & Bonuses
- Dispatch
- Construction & Commissioning
- Testing
- Metering
- Billing
- Force Majeure

Fig. 2 Typical items covered in a power purchase agreement (PPA).

Fuel supply agreement

- Term
- Quantity
  - Minimum
  - Maximum
- Fuel specifications
  - Primary
  - Back-up fuel
- Acceptance criteria
- Delivery point
- Metering
- Billing
- Force Majeure

Fig. 3 Typical items covered in a fuel supply agreement (FSA).

Lenders

Typically, nonrecourse financing is followed. Generally, the owners/investors contribute part of the total cost CAPEX (Capital Expenses) in the project, that is the investment, this could be of the order of 20%–30%, and hence the owners/investors raise the remainder through debt finance. Depending on how large the debt is, there may be a number of lenders, and there may be multiple debt tranches with different lenders and different conditions: term, interest rate, start of repayment, and fees. The lenders require a guarantee; hence, the project agreements are typically reviewed and will include appropriate clauses to protect the lenders' rights. The lenders request appropriate protection for their rights through the loan's life.

Advisers

In the various stages of the project, when the various stakeholders do not have the appropriate experiences in house, they engage advisers. These can be technical (e.g., engineering), environmental [for preparing or reviewing the environmental impact study (EIA)], or financial (including modeling and loan arrangers, legal, insurance, and tax advisers).

Contractors

Generally speaking, the project company deploys two main contractors: the engineering, procurement, and construction (EPC) contractors, and the operation and maintenance (O&M) contractors. The EPC contractor undertakes the execution of the project facilities per the EPC contract it enters into with the project company. Typically, the EPC contractor starts from site preparation. The EPC contract typically addresses the items covered in Fig. 4.

The project company deploys an O&M contractor, which handles the operation and maintenance of the plant upon its completion. Fig. 5 shows items typically covered under an O&M contract.

Payment Structure

Generally, an IPP entering into a PPA with an off taker under a LTPP scheme expects to get a reasonable return on its equity or an IRR (Internal Rate of Returns) for the project. Thus, the IPP weighs the expected revenue vs. the expenses through the life of the PPA or the term. The revenue side is established in the PPA and is generally broken into two types of charges: fixed charges and variable charges.

The fixed charges are in turn broken into two categories: capacity charge and O&M fixed charge. The capacity charge is meant to recover the capital costs initially invested in the project, including development costs, EPC costs, the

EPC contract

- Definitions
- Contractor responsibilities
- Owner responsibilities
- Notice to proceed
- Compensation & Payment
- Testing and acceptance
- Substantial completion and final acceptance
- Variations
- Warranties
- Remedies / Liquidated Damages
- Subcontractors
- Insurance
- Indemnity
- Assignment
- Termination & Suspension
- Force Majeure
- Confidentiality
- Site risk
- Attachments

Fig. 4 Typical items covered in an engineering, procurement, and construction (EPC) contract.

financing of the loans for the loans duration, and other applicable capitalization expenditures.

The fixed O&M charge relates to the fixed expenditures that the IPP encounters irrespective of how the plant was run (e.g., the cost of the O&M staff deployed to operate and maintain the plant on a continuous basis).

The capacity charge and the fixed O&M charge represent a payment stream that is paid by the off taker periodically (in most cases, monthly), provided that the IPP has what is termed declared capacity available for dispatch by the off taker. These charges are generally based on the declared capacity, in terms of kilowatt; thus,

O&M Contract

- Definitions
- Term
- Scope of services
- Responsibilities and Rights of owner
- Reporting and Audits
- Operating plan
- Maintenance program
- Expenses
- Subcontracting
- Fees, Bonuses and Penalties
- Remedies
- Force majeure
- Indemnity
- Limits of liability
- Insurance
- Confidentiality
- Assignment
- Dispute resolution
- Attachments

Fig. 5 Typical items covered in an operation and maintenance (O&M) contract.

the charge is specified in X \$/kW per month for the capacity charge and Y \$/kW per month for the fixed O&M charge.

On the other side, the variable charges—generally termed energy charges—relate to the net electrical energy taken by the off taker; thus, these charges are based on the electrical energy in kilowatt-hour per month.

These charges are broken into two types: fuel charges and O&M variable charges. The first charge relates to the costs of the fuel used to produce the electrical energy delivered by the IPP. The efficiency of the units in the plant, depending on the dispatch instruction subject to the technical specifications of the units, figures into the fuel consumption and, hence, the related compensation. The O&M variable charges are meant to address the expenses the IPP encounters as a result of running the units in the plant, such as the cost of consumables and maintenance related to wear and tear associated with the unit's operation. In some cases, the PPA takes the form of what is called an energy conversion agreement (ECA). Under such agreement, the off taker assumes the responsibility of the fuel. In this case, the fuel charges are not included in the off taker's payment.

The above payment structure gives the basic payment structure for a typical LTPP. In addition to the above, there may be ancillary payments/adjustments to account for, such as:

- Unit starts
- Shortfall in capacity/heat rate in tests compared with agreed specifications
- Shortfall in plant availability

In some PPAs, there may be allowances for bonuses if performance exceeds or falls short of the agreed-upon values.

For IPPs in countries other than the United States, the PPAs may use two components for these charges—one in U.S. dollars and another in the local currency.

Risks

The project agreements represent what the stakeholders have agreed upon in terms of the extent of the exposures that each party is willing to accept. The risk allocation for each project is quite complex. Exhibit 1 gives a simplified representation of the risk matrix from the perspective of the IPP and the off taker.

It is noted that the risk matrix takes into account the risks in each phase of the project (i.e., project development, plant construction, and plant operation). These phases are discussed in the following sections.

Referring to Exhibit 1, site conditions include but are not limited to archaeological material, buried objects, and soil characteristics. Environmental conditions include emissions, noise, and waste. Force majeure includes political and natural events.

As for the agreement on the tariff—i.e., the charges (referred to under the Payment Structure section above), escalation, and interest for delay in payment are generally agreed to.

Fuel-related risks—including quality, quantity, uninterrupted delivery, and backup fuel in the case of unavailability of the primary fuel—are of primary concern to the IPP. Fuel price is a major concern for the IPP.

In an international IPP setting, government guarantees are sought by the IPP for payments and other pertinent obligations for the off taker and the fuel supplier. In some instances, tax exemptions or holidays are offered by the host country. The currency exchange and local escalation are also addressed.

The lenders require many commitments to safeguard their interests until the loans are fully paid.

PROJECT DEVELOPMENT

For LTPP, the starting point may be a bid initiated by or on behalf of an off taker. Under such a case, the entity that wants to invite candidate bidders to undertake the project starts with steps to define the project and proceeds to qualify possible bidders and tenders the project. When the candidate bidders qualify and prepare their bids, they go through the development phases. This sequence is described below.

Project Definition

The entity that is undertaking the bidding on behalf of the off taker starts by evaluating the demand to establish the required capacity. It undertakes a feasibility study that compares options and alternative schemes for the project, including possible sites for the project. The feasibility study addresses the economics and the environmental impact. Based on the results of the feasibility study, the terms of reference for the project are prepared. These include the functional technical specifications. The site investigations are undertaken to complement the technical specifications. Then the request for proposal (RFP) is prepared, and the commercial conditions are spelled out. The RFP generally includes drafts for the project agreements. Further, the RFP addresses the evaluation criteria.

In parallel to the preparation of the RFP, the qualification of candidate bidders is completed.

When the RFP is issued to the qualified bidders, each of the candidate bidders initiates its development activities (refer to the next section).

Exhibit 1 Risk Matrix

Risk	IPP	EPC contractor	O&M contractor	Off taker	Fuel supplier
Site conditions	X	X		X	
Permits & consents	X	X	X	X	
Change in law	I	I	I	I	
Proven technology	I	O	I	I	
Force majeure	X	X	X	X	X
Political risk	I	I	I		
Insurance	X	X	X		
Plant construction					
Fuel terminal readiness		I			
Interconnection readiness		I		O	
Construction schedule	Delay liquidated damage	O	I		
Fuel quality	I	I			O
Plant	Performance	O	I		
Performance	liquidated				
Guarantees	damage				
- Capacity					
- Heat rate					
Environmental guarantees	Plant acceptance	O			
Payments	O	Perform in accordance with EPC contract			
Training		O	I		
Plant operation					
Availability	I		Perform in accordance with O&M contract	I	
Degradation	I		O	I	
Plant	I		O	I	
Performance					
- Capacity					
- Heat rate					
Fuel delivery	I		I	I	O
- Quantity					
- Quality					
Spare parts	O/I		O/I		
Payments	I			O	

X: Concerned with, O: Obligations, I: Impacted

Development Phases

The sponsors forming the group that is to bid for the project agree on the roles each group member is to undertake, and, on the sharing among them, this generally takes the form of a joint development agreement. The sharing of external costs for consultants retained by the group is addressed in this agreement.

The preparation of the bid in response to the RFP is the focus of the first phase of the development.

During this phase, the technical activities focus on the selection of the plant configuration and conclude with an EPC agreement with the EPC contractor.

The EPC agreement represents the basis on which the group bids in terms of the plant performance at site conditions and at different loads/dispatch. Technical considerations also cover the degradation in the performance with time. The technical considerations also extend to address the anticipated O&M. The technical data is then used as input to the financial model that the group formulates to simulate the finances of the project over the term.

The financial model is a key tool that the group uses to simulate the financing according to the equity sharing and the loan conditions. The financial model simulates the revenue streams and the expenses over the term of the project.

With the aid of the financial model, the group establishes the bidding tariff it is going to offer.

As part of the bid preparation, the group, with the aid of its advisers (consultants), establishes the loan agreements and the acceptable terms and conditions for its bid.

If the group's bid is selected, the second phase of the development proceeds. During this phase, negotiations of all the project agreements are completed, leading to initialing the project agreements.

In parallel, the group completes its negotiations with the EPC contractor and the O&M contractor. It also completes key steps with the lending institutions to reach agreement with the lenders for project financing.

With the initialization of the project agreements, the group completes the negotiation/finalization of the loan agreements and all necessary documentation to reach financial close. All activities to secure necessary permits and consents, including preparation of the EIA, are also completed.

In parallel, the group completes the formation of the project company. When financial close is reached, project implementation by the project company occurs as the IPP commences. Initial funding of the EPC contractor takes place; notice to proceed is given; and plant construction starts.

PLANT CONSTRUCTION

Upon financial close, the EPC contractor is given notice to proceed by the project company. The project execution commences. The EPC contractor proceeds with engineering. The project company may engage an owner's engineer to oversee the execution of the EPC contract. The EPC contractor submits a detailed schedule for the project. The various drawings and specifications are completed by the EPC contractor; they are reviewed and approved as applicable by the project company. Procurement of long-lead items is initiated, and site preparations get under way. Then construction proceeds with civil works, and the mechanical and electrical erection of equipment. Then comes commissioning. Finally, the testing, including reliability and performance tests, ends with the completion certificate. This marks the commercial operation date.

During this phase, meeting schedule milestones are monitored. Delay liquidated damages are effected as applicable, and everyone involved should avoid delays because of the economic and financial consequences. Equally important is meeting all performance guarantees specified in the EPC contract.

PLANT OPERATION

During the plant construction phase, the O&M contractor proceeds with various mobilization activities. These include hiring personnel; training; reviewing procedures and documentation; and receiving and storing spare parts, tools, consumables, etc.

From the start of commercial operation, the O&M staff takes charge of the operation and the maintenance. The operation means responding to the dispatch orders, monitoring, and reporting on plant performance. The maintenance includes the regular and planned maintenance, and responding to forced outages that may be encountered.

The O&M staff ensures conditions that attain maximum availability and optimum performance for the most economical operation under high reliability. The O&M staff also ensures the completion of all regular inspections and tests in accordance with operation manuals and the requirements of the project agreements. Ensuring the highest levels of safety and operating the plant in accordance with the environmental permits are essential. Testing for the verification of degradation also is performed per the PPA requirements.

CHANGING WORLD

Although the discussions presented in the previous sections give reasonable representation of many of the various aspects of the IPP world, the dynamics in world economy, and the oil and fuel markets, environmental concerns continue to affect the electricity industry.

Competitive forces in any market have major influence. The growth of major companies specializing in deregulated power generation and the entry of newcomers to the IPP world continue to be influenced by evaluating the individual opportunity and the risks involved. The expertise accrued and the previous experience/record enable adjustment to new conditions in a market the company is already in or a new market.

Changing regulations also has its impact. The perception of societies of deregulated power generation continues to be debated between advocates of privatization and proponents of public/regulated electricity.

Technology

The advances in gas turbine and combined cycle technology have propelled the growth of IPPs around the world, particularly with the increased use of natural gas to fuel many IPPs. Renewable IPPs are also gaining a larger share due to increased awareness of using cleaner power generation to protect the environment. As this trend continues to emerge, hydro and wind power are seeing steady growth.

CONCLUSIONS

In this entry, the history of how IPP came about was reviewed, and the stakeholders for a typical LTPP-IPP were presented. Through this discussion, the roles of the various participants were presented. Then the typical stages for an IPP project were discussed.

The discussions in the previous sections provide general information, yet this information should be a good basis to use in varied situations.

The IPP industry continues to grow with improvements in technology and the emergence of new technologies.

ACKNOWLEDGMENTS

The author wishes to acknowledge the very valuable information and knowledge acquired through participation in numerous IPPs/IWPPs while at Enron and since joining Consolidated Contractors International Company in 2001. Through extensive exposure to and interfacing with outstanding and experienced developers, advisers/consultants, original equipment manufacturers (OEM), and O&M principals as well as clients, the author has accumulated extensive experience. For the purpose of preparing this entry, the author opted to write it from this perspective rather than depending on citations. The author hopes that its contents will be useful to the reader, particularly to a novice interested in a comprehensive introduction to IPPs.

In the discussions of this entry, the author opted not to get into material that is covered in other sections of this encyclopedia. Further, the reader will find many sources of information on IPPs on the Internet.

The author also wishes to acknowledge the valuable input received from the reviewers of this entry.

Industrial Classification and Energy Efficiency☆

Asfaw Beyene
Department of Mechanical Engineering, San Diego State University, San Diego, California, U.S.A.

Abstract
In this article, energy consumption data for 300 manufacturing plants in Southern California are collected and analyzed by standard industrial classification (SIC) code. The results show that in order of magnitude, combined heat and power (CHP), variable speed drive (VSD), and compressed air systems offer the largest opportunities to reduce energy use in small and medium sized manufacturing plants. It also follows that the SIC is not a convenient tool to classify energy use, and in fact, to date there is no such convenient code to sort manufacturing plants by their energy intensity. The energy use profile cannot be systematically fitted to existing industry classification for the group of targeted plants. Opportunities and challenges related to capturing the potential energy savings in small and medium sized plants are also summarized.

BACKGROUND

The subjects of this article are data collected over several years through the Industrial Assessment Center (IAC) (The center is funded by the Office of Industrial Technology (OIT) in the U.S. Department of Energy (DOE), which is responsible for promoting and funding research for energy efficiency in industrial sectors.) at San Diego State University (SDSU) from 300 plants in 18 major standard industrial classification (SIC) (the North American Industrial Classification Standard (NAICS) has replaced the SIC code. Both classifications will be used in this article although most of the study was conducted for the latter) categories. It is believed that these data from diverse manufacturing plants can provide a generalized hint to better serve manufacturing and also develop better research strategies and funding policies. It can also be duplicated elsewhere as a useful measure to promote energy efficiency and link academia with industry. Although most energy efficiency strategies and technologies are applicable to manufacturing plants regardless of their sizes, there are issues unique to small and medium size plants such as those targeted by IACs that limit the abilities of these plants to identify and implement energy efficiency measures. This program is funded by the DOE to defeat these limitations.

Small and midsized plants are defined initially by the Organization for Economic Cooperation and Development (OECD), established on December 14, 1960, in Paris to promote sustainable economic growth. Members of OECD include Australia, Japan, U.S.A., Turkey, Canada, Greece, Iceland, New Zealand, and countries of Western Europe. The group formally associated sustainable growth with small and midsized plants and addressed the issue of charting industrial policy. The OECD committee addressed the diffusion of technology, and creation of new firms, supportive networks, financial incentives, and regulatory framework for target industries. Small and medium firms were recognized and encouraged as a component of sustainable growth. Department of Energy's support to such industry sectors stems from this early recognition of the role of these industry sectors in the larger scheme of economic growth.

ENERGY SAVING OPPORTUNITIES

Among the 300 plants analyzed, 19% are rubber and plastic manufacturing starting with SIC code of 30, whereas 15% are fabricated metal products starting with SIC code of 34. These two manufacturing groups, accounting for about one-thirds of the total, deal with some sort of molding process. The nature of the variable load in molding machines typically offers great opportunities for variable speed drive (VSD) application. Among the remaining plants surveyed by IAC/SDSU, electrical and electronic plants starting with SIC code of 36 comes third. Eleven percent of this group of plants was surveyed. Rubber and plastic plants are concentrated in Southern California as evidenced by the ratio of plants assessed, but the high number of metal manufacturing reflects more IAC's and DOE's emphasis on Industries Of the Future (IOF) (IOFs are defined by DOE, and the list can be viewed at DOE's web site.) in the late 1990s and early 2000s than the statistical concentration of the metal industry in Southern California. The plants analyzed here are mostly in the counties of San Diego, Orange,

☆ This entry originally appeared as "Energy Efficiency and Industrial Classification" in *Energy Engineering*, Vol. 102, No. 2, 2005. Reprinted with Permission from AEE/Fairmont press.

Keywords: Energy; Industry; SIC; NAICS; POEIC.

Table 1 Comparison of standard industrial classification (SIC) and north American industrial classification standard (NAICS) categories

Code	1987 SIC	1997 NAICS	
20	Food and kindred products	311	Food
21	Tobacco manufactures	312	Beverage and tobacco product
22	Textile mill products	311	Textile mills
23	Apparel and other textile products	314	Textile product mills
24	Lumber and wood products	315	Apparel
25	Furniture and fixtures	316	Leather and allied product
26	Paper and allied products	321	Wood product
27	Printing and publishing	322	Paper
28	Chemicals and allied products	323	Printing and related support activities
29	Petroleum and coal products	324	Petroleum and coal products
30	Rubber and miscellaneous plastic products	325	Chemical
31	Leather and leather products	326	Plastics and rubber products
32	Stone, clay, glass, and concrete products	327	Nonmetalic mineral product
33	Primary metal industries	331	Primary metal
34	Fabric metal products	332	Fabricated metal product
35	Industrial machinery and equipment	333	Machinery
36	Electrical and electronic equipment	334	Computer and electronic product
37	Transportation equipment	335	Electrical equipment, appliance, components
38	Instruments and related products	336	Transportation equipment
39	Miscellaneous manufacturing industries	337	Furniture and related product
		339	Miscellaneous

Indu–Inter

Riverside, and Los Angeles. The NAICS code has been suggested to replace the SIC code (Table 1).

The combined electric consumption of the 300 surveyed plants was 1318 GWh/year 14% of the plants used no natural gas or used a negligible amount. A total of 1996 energy conservation opportunities (ECOs) were recommended, where 61% of the plants did not offer natural gas savings. The remaining 39% made a combined total gas savings of 16,101,416 therms/year The total savings are about $19 million for electric and about $8.5 million for natural gas, which results in about $27.5 million for both, about 20% of the total energy use. The payback period is calculated as the total implementation cost divided by the total annual cost savings. A summary of the energy use profiles and costs of the assessed plants are shown in Table 2.

Plotting the annual total cost of energy ($/year) of the 300 plants as a function of their recommended annual cost savings establishes a somewhat linear relationship between the two parameters; the higher the energy use the higher the gross savings (Fig. 1). The linear graph further shows that a plant may save about $0.28 for every dollar spent on conservation. The coefficient of variation, $R2=0.59$ suggests about 60% confidence in a linear relationship between energy cost and saving potentials. However, higher savings are accompanied with higher implementation cost. On the average, about $1.36 will be spent for every dollar savings. With $R2=0.449$, the relationship between energy saving ($/year) and implementation cost has about 45% confidence to have a linear relationship (Fig. 2). The most common recommendations for energy saving strategies are shown in Table 3.

Quantitatively, lighting retrofits (30%) and air leak repair (14%) constitute the top two energy saving opportunities. Heating, ventilating, and air conditioning (HVAC) (3%) and fuel switch from electric to natural gas (3%) are the least recommended (Fig. 3). Combined heat and power (CHP), although only 10% by quantity recommended, offers the largest cost savings at 56% total. Consequently, although CHP is not frequently recommended, it provides the largest cost and energy savings (Fig. 4).

Based on the analyzed data, CHP offers the largest opportunity for cost saving. Fuel switch offers the second largest cost savings but often comes at the cost of increased energy use. Fuel switch entails replacing

Table 2 Energy use summary of surveyed plants

Total electric use	1,318 GWh/year
Total cost of electric energy	$107,481,270/year
Total electric energy savings	307,784 MWh/year
Total electric cost savings	≈$19 million
Average electric use	4,438 MWh/year
Average electric energy savings	1,036 MWh/year
Total natural gas use	71,767,530 therms/year
Total cost of natural gas	$33,092,324/year
Total natural gas savings	16,101,416 therms/year
Total natural gas cost savings	≈$8.5 million
Average natural gas use	241,642 therms/year
Average natural gas savings	54,214 therms/year
Total energy (electric and gas) cost	$140,573,594/year
Total cost savings (electric and gas)	≈$27.5 million
Total implementation cost (electric and gas)	$45,802,482
Average payback period	18 months

electricity with natural gas burners where the technology allows. Switching the energy source is known to save large amounts in Southern California because of the high cost of electricity and demand charges. This may continue to be the case unless the price of natural gas rises sharply relative to the price of electricity. Variable speed drives offer the third largest energy saving opportunity. Nine of the 20 largest energy saving opportunities in all the survey are related to CHP, eight to VSD, and three to compressor systems. Thus, excluding fuel switch, the largest opportunities for energy savings can be summarized as:

- Waste heat recovery and CHP
- Variable speed drive in metal industry and other variable load applications
- Compressor systems
- Modernizing aging equipment.

Waste Heat Recovery and CHPs

Combined heat and power applications constitute qualitatively the single most prominent opportunity for industrial energy savings, mainly because properly designed CHP systems offer superior efficiencies, at times exceeding 80%. Combined heat and power is so rewarding because the source of energy is often waste heat, and there would often be no net emission addition at the site.[1]

Among the 300 plants discussed here, almost all savings amounting to one million or more kilowatt per year come from CHPs. The few exceptions are VSDs on injection and die-cast processes when the plant in question has several large machines. The CHP installation also includes incineration exhaust heat recovery.

In spite of such great opportunities emanating from use of available waste heat, and in spite of the fact that CHP enjoys some level of support through state incentives, its implementation has been slow, even for cases where the savings are large and the payback is short. The market has not yet developed the necessary confidence, mainly due to regulatory and technical challenges. The main support for

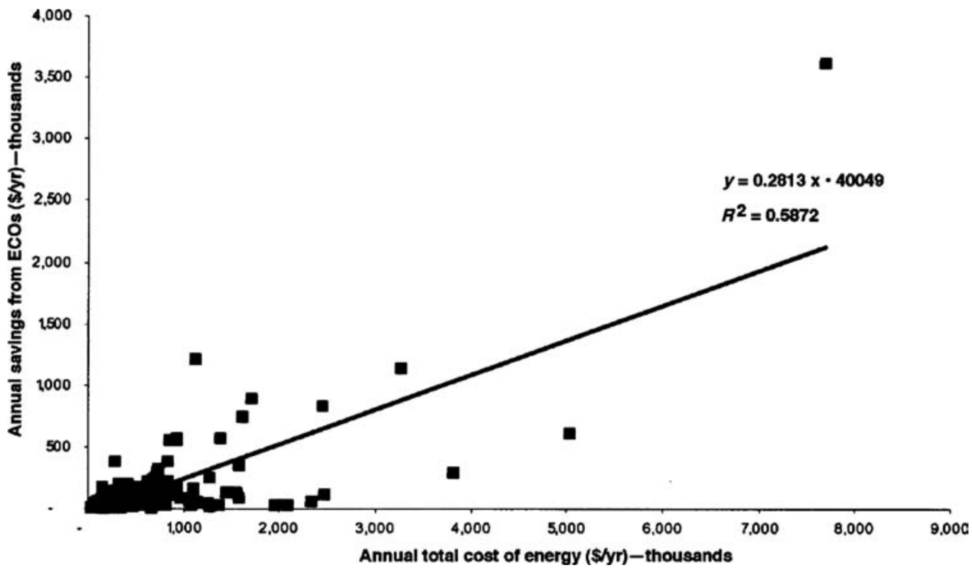


Fig. 1 Annual total cost of energy ($/year) vs their respective annual cost savings.

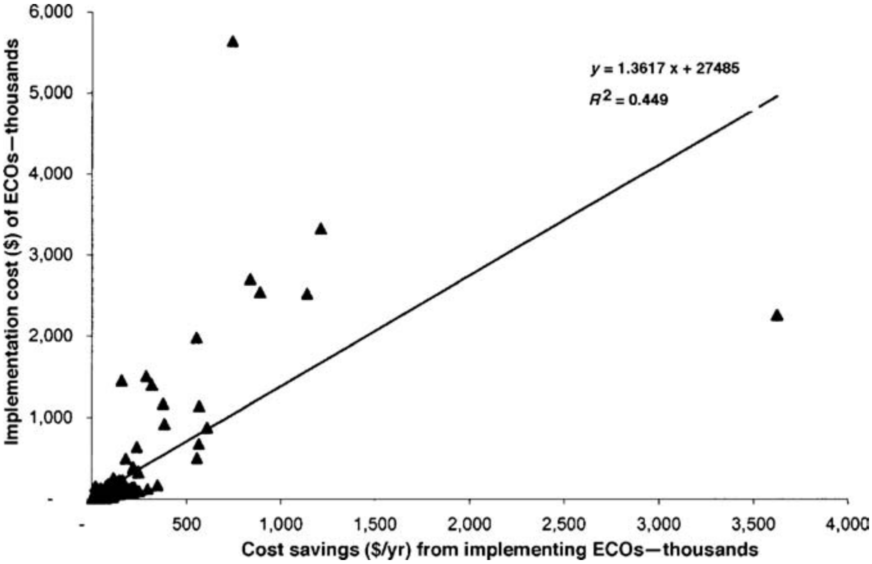


Fig. 2 Relationship between annual cost savings of energy conservation opportunities (ECOs) and their implementation costs.

Table 3 Common types of energy saving recommendations by category

Recommendations	Percentage of total recommendations	Percentage of total savings
Combined heat and power (CHP)	10	56
Switching energy source	3	25
Insulation	9	4
Power factor	12	4
Lighting	30	4
Air leaks	14	3
Other	19	2
Heating and cooling	3	<1
Total	100	100

CHPs in the United States started in 1978 when the U.S. Public Utilities Regulatory Policy Act (PURPA) was formulated, allowing excess energy produced by CHPs to be sold back to the main grid. These guidelines gave impetus to increased use of CHPs, which until then was restricted by rules that favored a tight monopoly in the power market. According to Energy Information Agency (EIA), by 1996, the United States had an estimated 51 GW of installed CHP capacity, about 6% of the total U.S. electric generation.[2] But unresolved issues related to technical and connection standards remained, and the growth of CHP was hampered. As a result, the 6% still remains lower than that of Europe, which has 9% CHP, and much lower than that of Denmark, at 40%. Estimates of future growth of CHP in the United States have been inconsistent. According to EIA,[2] at most a 6 GW increase can be expected by 2010. A more recent report suggests up to 31 GW of CHP additions by 2010.[3] Other estimates of potential for additional CHP capacity by 2010 vary from as low as 7.1 GW to as high as 60 GW.[4] Despite this disagreement on estimated growth, there is consensus on at least doubling the current CHP output by 2010. This is also consistent with DOE's challenge made to the CHP industry to double the U.S. CHP generation capacity by 2010.[5] But the inconsistency also points to a lack of accurate market data, especially on smaller CHP units, and the inability to anticipate regulatory and policy trends.

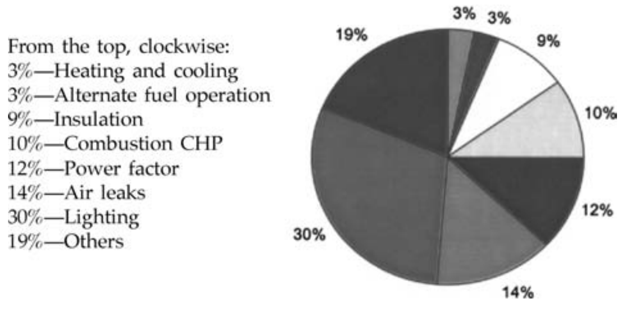


Fig. 3 Graphical representation of energy saving recommendations by category.

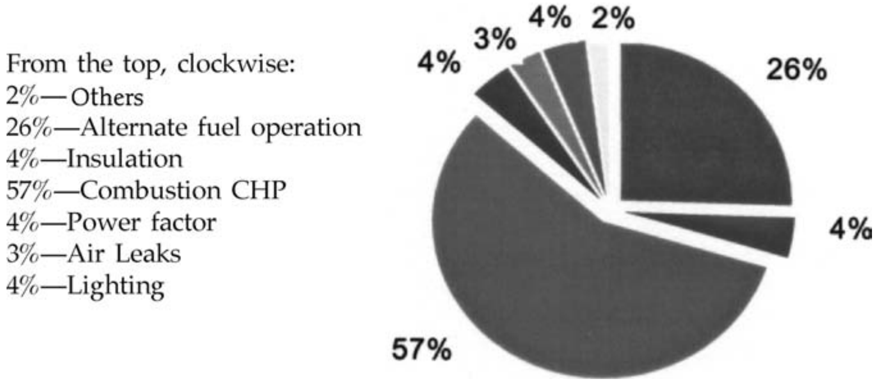


Fig. 4 Graphical representation of energy saving recommendations by cost savings.

Energy Information Agency conducts a survey of nonutility power generators on a yearly basis to collect data on CHPs exceeding 1 MW. Other data collected by DOE and independent research institutes are sporadic, fragmental, and limited in their outreach.

The most difficult regulatory hurdle is probably the new source review (NSR) and the prevention of significant deterioration (PSD), particularly in nonattainment areas, cities where pollutant levels exceed the national air quality standard. The NSR requires that facilities obtain a permit before beginning construction of a new facility or before significantly increasing emissions. In nonattainment areas, the installation of a new CHP, even if it is proven to reduce global emissions, may still be in violation of PSD. Outside nonattainment areas, new installation will mean new sources of emission, adding to existing emission levels, but the new sources would still be much lower than what would have been emitted from the electricity supplying power plant. Since electricity is usually distributed from central plants over long distances, a CHP may increase on-site emissions at the local level, and therefore trigger NSR even if it displaces emissions from the remote fossil fuel combustion power plant that supplies electricity to the facility. This means an overall reduction in emissions as a result of new CHP installation is not only ignored in assessing global pollutions, but locally it also triggers expensive and time-consuming regulatory processes. The lack of clear understanding of such complex rules has more often forced plant managers to take a safe route, which is to avoid facing regulatory bureaucracy by avoiding CHPs altogether.

Interconnection standards for back-up capacity from utility and off-site sale of excess electricity is locally regulated or left for negotiations, at times resulting in unfair market competition that suppresses CHPs. There is no single and simple stream of process to ease and simplify existing complex permit procedures well-discussed by Kolanowski.[6] Even systems with very high efficiencies have no blanket regulatory support.

In the United States, federal support for CHP has been increasing in recent years. One such major development is the 2004 report of the National Energy Policy Development (NEPD) Group. The group recommended that the president direct:

- The secretary of the treasury to work with congress to encourage increased energy efficiency through CHP projects by shortening the depreciation life for CHP projects or providing investment tax credits.
- The Environmental Protection Agency (EPA) to work with local and state governments to promote the use of well-designed CHP and other clean power generation at brownfields sites, consistent with the local communities' interests. Environmental Protection Agency will also work to clarify liability issues if they are raised at a particular site.
- The EPA administrator to promote CHP through flexibility in environmental permitting.
- The administrator of EPA to issue guidance to encourage the development of well-designed CHP units that are both highly efficient and have low emissions. The goal of this guidance would be to shorten the time needed to obtain each permit, provide certainty to industry by ensuring consistent implementation across the country, and encourage the use of these cleaner, more efficient technologies.

These recommendations have been praised for legitimately embracing the CHP technology. They are important steps towards rewarding high efficiency systems with investment credits, tax incentives, exemption from fees, shorter depreciation tax life, and blanket regulatory support. But critics also say they have not gone far enough to establish parallels between efficiency and pollution reduction by, for example, linking emission limits to the amount of net power output from the plants, not to the amount of fuel required to run a generating plant. In the current scenario, less efficient plants requiring more fuel input will be allowed more polluting emissions than efficient plants requiring less fuel for the same power output. However, there is a general consensus that the NEPD recommendations are necessary steps to advocate and promote the environmental benefits of CHPs. The impact of these recommendations on the overall progress of the CHP market and technology remains to be seen. To foster CHP, interconnection standards must be better regulated and a single and simple stream of process to ease and simplify existing complex permit procedures should be established.

VSD for Variable Load Applications

Synchronous motors are powered from a fixed-frequency electrical source, i.e., they turn at a fixed rounds per minute (RPM). If such machines always operated at full capacity there should be no reason to employ VSDs. However, they often run at part-loads, giving an opportunity to save energy with reduced speed when the load drops. This is possible due to advances in the electronics industry that now offer accurate and reliable breeds of new motor control devices. The a.c. inverter or VSD is placed between the motor and its power source to collect a signal indicating the load of the motor. The signal is then used to vary the frequency of a.c. electrical power to match the load, which results in power savings by running the motor at all part-load speeds without significant penalty to the efficiency.

Variable speed drives are commonly used in injection molding machines, HVAC systems, conveyers, packaging lines, and many other processes with variable load. For injection molding machines, research has shown that 20%–40% energy savings are possible. A field test conducted by the author in air-handler fan applications has also confirmed savings of the same magnitude.[7] The 30% range saving is generally applicable to cycling machines with similar intervals to injection machines. The idea is also valid for mixers with varying density and viscosity of the mixed medium, an application still requiring research. Mixing is common in paper and cellulose byproducts, chemical, and food industries. The amount of energy depends on such mixing characteristics as the density, viscosity, batch, composition, and required consistency of the final mix. This is also the case in paint mixing and liquid food processing such as ice cream and soup. Since one or more of characteristics of the mixture almost always varies in time, it is advisable to vary the input energy to match it to the load, and as a result, reduce energy use.

A large percentage of the mixing energy in these processes is dissipated as heat, usually requiring cooling of the mixing tank. This may require additional chilled water or tower cooling with pumps and fans, involving other lateral capital and maintenance costs. This means, in addition to energy savings anticipated as a result of the matched energy input to the actual variable load, VSDs have several other related applications. Reducing the input energy also means reducing the generated heat, and therefore, less energy load on the cooling system.

Advantages of VSDs can be more than simple savings, and may include reduced utility peak demand charges, a lessened space cooling load required to remove the energy dissipated in the form of heat, less noise, and reduced tear and wear problems on motors and other parts. The ability of digitally administering the RPM can also improve product quality by allowing more accurate control of the process. It can save time by better synchronizing the starting and ending time, and possibly even using the change in RPM at constant input power to signal the change in viscosity or other fluid properties.

Compressor Systems

Compressed air systems offer the third largest energy saving opportunity, but generally have the quickest payback period. In the United States, compressors alone use 18% of plant energy consumption.[8] Virtually all manufacturing plants assessed have compressors, and almost all compressed air systems had leaks. It can be argued that air leaks are the most common waste of energy in compressed air systems. In many cases, compressed air is supplied at much higher pressure to be throttled down to the required set point. Some plants need different levels of pressures in different parts of the plant, forcing the compressor to work for the highest pressure output needed by one or few machines. The following is a list of the most common sources of energy waste in compressed air system.

- Air leaks.
- Over-compression.
- Running the compressor when not needed, time-clocks to shut off at night and weekends.
- Expensive condenser cooling.
- Oversized compressors.
- Wrong application, such as using compressed air where blowers could do the job.
- Mismatch of process and compressor hours, for example, solenoid valve could be used to shut off compressed air supply from machines that need it only when operating.
- Cooler outside air intake.
- Use of compressor waste heat.

Measures that could reduce or eliminate energy waste in compressed air systems are typically less complicated and inexpensive. Air leaks can be repaired; compression can be reduced to match the site need, and in cases where the pressure demand varies, multiple compressors can be installed to isolate and serve the peak areas with dedicated compressors. Because of the prevalent use and associated presence of energy waste so common to air compressors, more serious regulatory and research attention to these systems would be productive.

Modernizing Outdated Equipment

The typical small or medium-sized plant has no in-house expertise for energy management. Such a plant, therefore, has to rely on expensive audits which can be biased if the vendors have a motive and market interest in their ideas. More aggressive training programs and workshops can help in technology awareness and transfer of efficient technologies to the manufacturing floor. More creative ways of distributing incentives to aggressively promote replacement of aging and inefficient equipment will also help small and medium sized manufacturing plants that often lack funding to modernize their plants. Low interest rate financing for proven technologies or concepts and promotion of not-for-profit institutes that promote efficiency will help build a good case for energy savings.

Visits to hundreds of manufacturing plants proved that equipment in old neighborhoods is also relatively old, especially noticeable in large and capital intensive blocks such as the metal industry, which are also relatively expensive to retrofit. The forging and casting industries are very energy intensive, with obsolete ovens and heat transfer control systems in most cases. This will be discussed, taking aluminum forging as an example.

Of the several process variables controlling closed-die aluminum forging deformation characteristics, die temperature is widely recognized as the most critical. For the majority of aluminum alloys forged, optimal die temperature for die lubricant performance and work-piece deformation characteristics is around 700°F. Consequently, on-press die heating is an essential thermal process for successful, cost efficient, and high quality aluminum forging manufacture. However, current on-press die heating systems are inefficient, with poor control systems. Process parameters, such as air–gas mixture and volumes are typically manually controlled. There is no research-driven knowledge base and understanding of the on-press die heating process. Most of the technologies currently in use, including burner, air–gas mixture systems, and controls, are over 30 years old.

Billet heating is the second largest energy user, as well as the second most important forging process variable controlling aluminum forging deformation. Required input billet temperatures are in the range of 700°F–850°F. Commonly, billet heating is accomplished by direct-fired natural gas conveyor ovens, which suffer from large thermal losses. Forging stock is charged in one end and exits the other through openings or open doors. Furnace and combustion systems have burner blowers with heat input ducted directly into the working zone.

Solution heat treatment (SHT) and aging thermal processes for aluminum alloy forgings impart the final mechanical properties at 890°F–995°F to achieve metallurgical objectives, and therefore, are very critical processes in the manufacturing flow. Because of the high-SHT temperatures, aluminum's affinity for hydrogen and rapid quench requirements, SHT furnaces are indirect fired (radiant tube), natural gas drop bottom batch furnaces. Required SHT practices for aluminum alloys necessitate precise control of the work piece temperature (±10°F or better) and time-at-temperature.

Artificial aging of aluminum alloys is a single- or multiple-step thermal process conducted at temperatures in the range of 200°F–400°F for times from 10 to 24 h Furnaces employed for artificial aging are car bottom, batch, and natural gas direct-fired aging types. Aging also requires precise control of the work piece temperature (±10°F or better) and time-at-temperature.

Sometimes dies are preheated prior to installation in the press in car bottom natural-gas fired ovens to temperatures of 600°F–800°F. The ovens are constantly opened, exposing the load to ambient temperature with attendant thermal losses. Research would help investigate the use of multicompartment ovens to reduce thermal losses. There remains a lack of fundamental understanding of the optimum and most efficient technology for die preheating that will support a large number of presses. A significant opportunity exists to effect reductions in energy use through capture of existing energy conservation or alternate heating technologies. Some of these energy saving opportunities include [9]:

- die and billet coatings to enhance heat transfer
- flat flame and/or rotary burner systems
- electric and/or gas infrared high energy flux heating systems
- state-of-the-art furnace burner and combustion systems
- high velocity recirculating air gas-fired furnace systems
- closed-loop and microprocessor control of thermal processes

Industry Classification and Energy Savings

To assess the relationship between energy consumption and industrial classification for small and medium sized plants, the following parameters were further calculated for each SIC code of the assessed plants:

- Gross annual sales ($/year).
- Number of employees.
- Electrical energy consumption (kilowatt per year).
- Gas usage (therms/year).
- Electric rates paid ($/kilowatt hour).
- Gas rates paid ($/therms).
- Energy conservation opportunities savings ($/year).
- Energy conservation opportunities implementation costs ($).
- Types of recommended energy saving opportunities.
- Energy savings (kilowatt hour and therms).
- List of energy consumed by services.
- Energy saving recommendations implemented.
- The total energy consumed per employee.

- The gross annual sales per plant annual energy consumption.
- The total annual sales per employee.

These values were calculated and graphed as a function of the SIC code. In addition, the last three values were plotted for four SIC codes of 3089, 3672, 3679, and 3728, which are the four most frequently surveyed plants by SDSU/IAC. The graphs showed that they vary drastically for each SIC group. Only a few plants have fairly consistent energy use per employee. Standard industrial classification 33 has somewhat consistent annual sales per energy use. Other plants have striking deviations suggesting gross variations in energy intensity even within the same SIC classification. For instance, the three parameters for 25 plants with SIC 3089 have gross annual sale/MMBtu (1 MMBtu = 0.293 MWh) varying from as low as about 500 to as high as about 4500.

All calculated values proved that the industry classification standards do not reflect the energy consumption, i.e., SIC and NAICS offer little hint to make pertinent conclusion regarding the energy consumption profile for any group of manufacturing plants defined by these codes. Graphs plotted with SIC codes on the *X*-axis and other relevant parameters such as kilowatt hour, cost savings, implementation cost, etc. on the *Y*-axis showed no readable uniformity per SIC or NAICS code and such graphs offered no unique features that can be generalized for each SIC or NAICS group. Thus, the SIC may be very useful for industry manufacturing classification by product output, but it does not mark or represent any energy use profile for the coded group in the manufacturing class. Even within one SIC group, the energy intensity is drastically different depending on the building type, equipment age, personnel awareness, utility incentives, regulatory codes, etc. For example, more SIC 3089 plants have been surveyed by SDSU/IAC than any other plant. These plants, however, have drastically different levels of energy and cost intensity for virtually the same product. Simply stated, no energy parameter is consistent per SIC, even for plants manufacturing the same final products. This can be explained by inconsistencies in management skills and tolerance to aging equipment of the manufacturing floor. Furthermore, the industry concentration in California and any other part of the U.S. is uneven, with some specific process types more concentrated than others.

The IAC data are not even by region and industry type, and this probably impacts the quality of the analyses. For example, miscellaneous plastics (SIC 30), fabricated metal products (SIC 34), and electrical–electronic equipment (SIC 36) constitute about one thirds of the plants surveyed. This ratio will certainly vary in other regions of California or the United States.

Heat recovery depends on process, not output classification. Processes such as melting, incineration, heat treating, heating, and drying often have high stack temperature, i.e., there is a possibility to recover large amounts of waste heat.

CHALLENGES

Industrial Assessment Center has extensive experience in recommending energy efficient devices and measures that save millions of kilowatt hour per year. Most, if not all, such recommendations are mature technologies that have been in the market for years. And yet, IAC reports are received as a promotion of new technologies and concepts. This is because of poor technology transfer, particularly in small and medium sized plants that do not have in-house expertise to adopt efficient and innovative energy management systems.

The relatively low price of electricity and the general industry understanding of energy expenses as a small part of the production agenda that has to be paid without much of a fuss has led to a general acceptance of energy costs as an essential expense, and efficiency as an extravagance to be thought of at times of energy crisis or diminishing profit. This is a serious problem to promoting energy efficiency. During the 2001 energy crisis, the IAC received several phone calls from industries looking for ways to reduce the cost of energy. The voluntary calls completely disappeared by the winter of the same year when talk of the crisis receded. During the crisis, reaction to IAC offers was almost immediate, much faster than later or during the years that preceded the crisis. The conclusion is obvious: a rising cost of energy critically narrows the profit margin to a level where energy cost becomes a serious matter. A 15% profit for a plant with $30 million total annual sales results in $450,000 in net profit. Paying $100,000/year for utility bills still would leave $350,000 profit. It can easily be seen why the plant cannot tolerate a doubling or tripling energy cost.

Industrial Assessment Center data suggest that unless an idea represents large cost savings of, say, over $100,000, industry reaction to ideas of payback in more than 2 years is cautiously enthusiastic or casual. Ideas that are commonly known, such as T-8 lighting, or ones with which the plant personnel are very familiar, have better chance of implementation. This explains why the implementation rate for great conservation opportunities such as CHPs offering millions of dollar savings is generally not as high as expected.

Lack of monitoring and recording energy consumption of individual processes is another major problem. For example, there is no tradition of kilowatt per ton monitoring of large chiller systems, which means ongoing energy efficiency of such systems, once installed, is not known. Few plants have consistent monitoring of air/fuel ratios for combustion, and air flow rates of high-temperature ovens have to be estimated from oven design capacity, since flow rates are rarely monitored. Without fairly accurate knowledge of the airflow rate, it is not possible, for example, to size the heat exchanger for heat recovery.

Another area of challenge is the complexity of control systems. For example, VSDs offer significant savings in areas where the load varies considerably, by matching the energy input to the load instead of supplying a constant amount of energy above the maximum point of the cycle. However, VSDs require thorough expertise in selecting the signal input to the device, protecting the system from disturbing signals, and also assuring a clean environment for its operation. This is overwhelming for most small and medium sized plants that have no on-site expertise. Variable speed drive vendors do offer expertise to install the VSD in most cases, but plant engineers in general tend to be uncomfortable using a device whose internal working mechanisms they don't understand and whose service is totally independent of their employees; i.e., production may be at the mercy of an external expertise who may not be available when called for.

Limited application of some energy saving devices is also a problem. This is true for lighting where, for example, reflectors and skylights may not be useful in a dusty manufacturing environment. Retrofitting 12-ft lamps with T-8 is almost always too expensive for buildings that are not lit for long hours. Caution must be taken when recommending VSDs to a very dusty and oily environment, unless the VSD can be placed in a suitable environment and a remote control system can be afforded. One common problem of energy use in industry is the selection of either the wrong device or the wrong size for the same job. An example is the use of compressed air as an agitator, as a blower, or for low-pressure control systems. These are all cases where a fan can be used. Plants also use large compressors for small operations during the weekends and third shifts even though the plant is mostly closed during these hours. A proper measure would be to purchase a small compressor for small operations and also to serve as a lead unit where only part of the manufacturing is running.

Implementation of CHP systems involves complex engineering and permit processes to account for the several variables impacting the system sizing and equipment selection. Such variables include the site energy use, demand charges, excess energy price, and steam need. To date, there has been an attempt to enlist detailed input data required for a typical CHP plant.[6] A CHP mathematical model has also been suggested to calculate investment returns.[10] A simple, free, web-based and user-friendly procedure to assist in feasibility analysis of CHP sizing and selection has also been developed.[11] However, there are still interconnection issues and regulatory obstacles to CHP implementation such as the NSR and the PSD, particularly in nonattainment areas. Under the NSR, facilities are required to obtain a permit before beginning construction of a new facility or significantly increasing emissions. This means overall reduction in emissions as a result of new CHP installation triggers expensive and time-consuming regulatory processes. Interconnection standards and existing complex permit procedures[6] remain as obstacles to CHP. The lack of clear understanding of such complex rules has more often forced plant managers to take a safe route, which is to avoid facing regulatory bureaucracy by avoiding CHPs altogether.

The parallel between efficiency and pollution reduction is not supported by the current U.S. emissions standards because emission limits are based on the amount of fuel required to run a generating plant, not on the amount of net power output from the plant. Less efficient plants requiring more fuel input will be allowed more polluting emissions than more efficient plants requiring less fuel for the same power output. A more equitable emission calculation equation rewarding high efficiency has been discussed.[11]

For lighting, replacing T-8 offers an opportunity to replace ballasts and insert reflectors which could cut as much as 50% of the light energy consumption. Reflectors may not offer large enough savings to be undertaken as a separate measure. The implementation cost is much more affordable if replacing bulbs, ballasts, and inserting reflectors are conducted all at once. Some plants stagger this process, ending up paying a high-implementation cost.

Other challenges include:

- Long payback period caused by short operating hours, resulting in low kilowatt hour.
- Unfavorable past experience with a new technology, such as harmonic feedback in VSDs.
- Lack of awareness and time, inability to reach decision makers in the plant.
- Inaccurate data reading resulting in exaggerated savings or undermined implementation cost.
- Lack of awareness about various opportunities, such as availability of incentives, existence of new technology.
- Sizing and purchasing errors resulting in hiring unqualified or inexperienced contractors.
- Code limitations such as the number of skylights on a roof.

CONCLUSION

The IAC experience suggests that there is a tremendous opportunity to recover the large amount of industrial waste heat lost through stacks and exhaust systems. Even low-grade heat, such as from compressor motor cooling, can be used for various applications such as space heating or sludge drying. Innovative technologies, such as oscillating combustion, control system, and new alternative energy sources can be promoted with energy efficiency, eliminating the need for new fossil-fuel power plants. Legislation can be designed to better serve energy efficiency by limiting maximum stack temperature. This will "force" industry to reclaim high-temperature waste heat and also eliminate the hazardous nature of such high temperature exhaust. More research on VSD applications in other areas can also expand its applications. Regulatory supports like emission credits for all documented energy savings or developing a methodology to assess and quantify the environmental impact of energy saving measures in terms of NO_x and CO_2 reductions can also be very useful to reduce the energy intensity of manufacturing plants.

The SIC classification may be useful to compare products and industry profiling; it is less than useful for energy analysis. This is because energy intensity is more a function of the process, not the industry type. Such industry process, for example, includes melting, heat-treating, forging, etc. These are energy intensive processes that may or may not be present in the given SIC code.

ACKNOWLEDGMENTS

The author greatly appreciates Department of Energy's Office of Industrial Technology for the several years of funding granted to the Industrial Assessment Center at San Diego State University. Great thanks also go to the California Energy Commission and Mr. Pramod Kulkarni specifically, for supporting the data analysis.

REFERENCES

1. Laitner, S. *Comparison of Waste Heat Generation in U.S. Electricity Generation with Total Energy Consumption in Selected Countries and World Regions*; Office of Atmospheric Programs: Washington, DC, 1998.
2. Energy Information Administration. *Annual Energy Outlook 1999 with Projections Through 2020*, DOE/EIA-0383 (99); EIA: Washington, DC, 1998.
3. U.S. Environmental Protection Agency. Report prepared for EPA by ICF Kaiser, 9300 Lee Highway, Fairfax, VA 22031-1207, Telephone: (703) 934-3600, CHP Potential in the Emerging Competitive Markets Preliminary Base Case Results, 1999.
4. Bernow et al. *Policies and Measures to Reduce* CO_2 *Emissions in the U.S.*, 1998. Available at: http://www.tellus.org/energy/publications/policies&measures.pdf (accessed February 6, 2007).
5. U.S. Department of Energy. *Energy Department Initiative Aims at Recovering Industry Power Losses*, PRL-98-046 [press release]; DOE: Washington, DC, 1998.
6. Kolanowski, B. *Small Scale Cogeneration Handbook*; The Fairmont Press, Inc.: Liburn, GA, 2000; 41.
7. Beyene, A. *Variable Speed Drive Harmonic Test*. San Diego Gas and Electric, [unpublished report, 1996].
8. Malinowski, J. Using ASDs with variable torque applications, *Energy Matters (the Best Practices' quarterly of DOE's Industrial Technologies Program)*, March/April 2000. http://www.eere.energy.gov/industry/bestpractices/energymatters/full_issue.cfm/volume=10#a160 (accessed December 4, 2006).
9. Department of Energy, Office of Industrial Technology web site: http://www.eere.energy.gov/industry/aluminum/analysis.html (accessed February 6, 2007).
10. EEBPP *How to Appraise CHP: A Simple Investment Technology*; U.K. Department of the Environment: London, 1997.
11. Beyene, A. *Combined Heat and Power Sizing Methodology*, Proceedings of ASME, GT-2002-30567. Turbo Expo 2002, Amsterdam, The Netherlands, 2002.

Industrial Energy Management: Global Trends

Cristián Cárdenas-Lailhacar
Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida, U.S.A.

Abstract
This entry details energy usage and savings in industry are areas of interest, particularly for large energy users. This interest becomes more important when the associated costs of energy rise at values larger than expected, making it a concern for all size and types of facilities. Therefore, the goal becomes to reduce energy costs while keeping or improving the quality of the manufactured goods, which is not an easy task.

In light of the energy and environmental problems that our industrial society is facing today, energy conservation and environmental protection through good energy management (EM) practices appear to be a natural path. Recently, great effort has been devoted to developing computational tools that can help energy managers save energy. In this regard, the U.S. Department of Energy (DOE) has devoted great efforts to developing a series of programs, such as the Best Practices program, the Federal Energy Management Program (FEMP), and the Industrial Assessment Center (IAC) program. Support for industries and federal facilities (like FEMP) and some software developed by the best practices (BP) and IAC programs are discussed here. Others have recently been created by the author and his collaborators. The use of these packages can greatly improve your chances of exploring common and new energy-saving opportunities. Numerical examples of their use are given, and their features are discussed.

Many publications based on case studies on energy efficiency, recommendations, new technology that consider reductions of energy consumption, electric regulatory issues, energy forecasting, and energy purchase exist. With all of this information, it is hard for industries to decide what steps to take or what strategy to follow. Saving energy while maintaining product quality will have a significant impact on their bottom line. In this framework, this entry will show current and new trends for energy savings, encompassing an energy assessment of the facility that cannot be conceived without an energy balance (EB) and its added value.

The importance of the consequences of energy efficiency and its environmental impact are discussed in this entry. Current procedures, techniques, and results through a case study, the introduction of the concept of energy descriptors (a mathematical representation of the energy consumption of equipment that are major energy users), which passes through an inevitable EB, are analyzed. We will show that there is a correlation between energy consumption and potential energy savings and implementation that is not only intuitive, but can be quantified.

In this entry, we discuss the energy–environment relationship from a global (worldwide) and local perspective in terms of its importance in the development of countries and personal welfare. As we move toward the local influence, we discuss these effects through a case study by reviewing the current energy usage patterns of a number of Florida citrus juice facilities.

INTRODUCTION

In today's energy-addicted world, the efficient use of energy, savings, and research on new fuels and their use is becoming a daily concern almost all over the world. As discussed in other entries, countries rich in nonrenewable energy resources today should administer them wisely, particularly those coming from fossil fuels. Nowadays, every country is affected by energy issues. Our industrialized world requires more and more energy every day, which, among its many consequences, means more environmental problems.

Keywords: Energy balance; Energy assessment; Energy and environment; Energy and development; Energy descriptors; Energy software; Energy analysis; Alternative fuels; Symbiotic energy.

At the governmental level, the responsibility for the way and fashion in which energy is generated and consumed is tremendous, particularly from the most industrialized countries. The provision of policies and regulations that should be in place as national energy management (EM) policies are of primary importance (rates, taxes, incentives, etc.). However, governments are not the only players in the energy efficiency and conservation endeavor.

Facilities (industries, buildings, commercials, etc.) that usually were not concerned about their energy consumption now pay close attention and dedicate time and effort to reduce it. In the United States, federal buildings are not exempt; Executive Order 13,123 mandates that by year 2005, energy consumption should be reduced by 20% as of that one consumed in 1990, and by 25% by 2010. In the chain of energy usage, energy savings are affecting our lives and our environment. Specifically, each time that energy (or water) is consumed, some fuel has been burned, and some pollutants are released into the atmosphere. These pollutants, whatever their origin and concentration, are moved around the globe through the atmosphere via jet streams that do not exclude any countries or people; rather, they affect the planet as a whole. For some years now, research has shown an increase in the amount of acid rain, and the effect of some pollutants in producing depletion (a hole) in the ozone layer, which protects us from being burned by ultraviolet rays coming from the sun. The increase in pollution can have catastrophic consequences, and it is a continuous threat to our presence on the planet.[1]

Saving energy has additional benefits. It certainly allows for economic competitiveness as production costs are reduced, along with the intensity in industrial energy. Moreover, energy efficiency allows for fuel independence by reducing oil imports, which, in turn, reduces our vulnerability to oil embargos. In addition, it reduces the negative balance of trade as imported oil bills are also reduced, directly affecting payment balances.

Although energy has to be consumed in order to produce the goods and services that we need, we can certainly imagine that facilities (manufacturing, buildings, hospitals, etc.) can continue their process with less, more efficient-energy consumption through the implementation of EM programs. This is the use of engineering and economic principles to control the cost of energy so as to provide needed services in buildings and industries and reduce energy costs and pollution. Among the rainbow of proactive facilities for energy savings, two extreme types can be considered, and they do exist, with the majority sitting in between. On one end is the somehow appealing choice for some companies to do nothing about it; their reasons are that they have no time (they are too busy), no qualified in-house personnel to undertake an energy program/audit, no funds available to hire the services of an external consultant, or that energy is a minor fraction of their operations cost, and, of course, the golden rule that "If it's not broken, then don't fix it!" The other end of the spectrum is represented by well-known companies around the world that have made tremendous changes in their energy consumption with teams of energy managers. 3M and Walt Disney World are just two examples.[2] These teams are constantly performing internal energy assessments that recommend and implement energy conservation opportunities, allowing them to continue manufacturing their goods and providing services with the same high quality, but optimum energy consumption; that is, they use their energy in the most efficient manner possible. This translates to resource savings that positively impact the bottom line of these companies. Needless to say, the savings obtained by implementing the ideas recommended by these EM teams have greatly exceeded their salaries (i.e., "Work smart, but not hard"). Unfortunately, this second type of facility is not very common; however, their example should be our ultimate goal. What we learn from this is that those facilities employing an energy engineer or manager certainly reduce their production costs and are more successful. We shall return to this topic in the next section.

The way to achieve this goal does not seem elementary, as much needs to be accomplished in new technology through research and development. Partnership between academia, the government, and facilities (industrial or federal) is very appealing. In this regard, the U.S. DOE has devoted an enormous amount of effort to help industries and federal buildings, as well to reduce energy consumption. The Office of Industrial Technologies (OIT)[3] and the Office of Energy Efficiency and Renewable Energy,[4] with programs like the Best Practices (BP) program, the IAC program, and the FEMP,[4] are examples of how the U.S. government is becoming more and more a partner in energy conservation.

A case study will be presented in this entry to show current energy (heat and power) usage patterns' associated costs, energy indicators, and previously recommended energy savings measures. We propose a set of energy savings measures outlining some energy indicator values and benchmarks, which will reduce production costs ($/unit). Finally, key energy end uses and innovations are explored.

In the upcoming sections, we also discuss the effect of energy use on the environment and development of affected countries and their industries. Within this framework, we will discuss the importance of an EM program from the financial standpoint. This will be discussed in more detail through the use of an industrial case study, where we address current and new trends in energy auditing. Later, we discuss the use of current and newly developed software, and address alternative sources of energy in a symbiotic fashion. Finally, we draw some conclusions regarding the importance of energy savings and environmental conservation.

ENERGY, ENVIRONMENT, AND DEVELOPMENT

Per our previous discussion, it is easy to see that hiring an energy manager will probably not happen in many industries unless required by law. In other words, there are a few countries whose governments have understood the importance of energy consumption efficiency and reduction, and that this is not a government problem only, but everyone's problem. Hence, they require that, by law, every facility that manufactures a product must have an energy engineer/manager in the plant as part of their permanent staff. An example is Japan, whose manufacturing industry has reached impressive international levels of excellence.

Some countries in Latin America and the Caribbean, like Chile, are taking similar steps through their national energy agencies, as is their Comision Nacional de Energia (CNE), and its Ministry of Energy and Mines, which promotes energy savings and incentives to install cogeneration systems, for example. The same is true for Colombia. Other nations in this region, like Ecuador, promote energy research and offer companies the option of donating 25% of their taxes on power to institutions (including educational and research institutions), with obvious beneficial consequences. Peru offers tax reduction incentives for international companies that want to establish operations in the country, etc. At a local level, in states like Florida, for example, there are tax incentive programs for discounts on up to 100% of taxes paid on electricity.[5] All of these incentives have tremendous impacts on the economy of local industry and the country itself.

The economic costs associated with implementing environmental protection ranges from the conclusions drawn by scholars such as Porter,[6] who believe that environmental standards will increase business competitiveness, to scholars who find no empirical evidence of this through statistical, economic analysis. The Kuznets Curve is an inverted-U relationship that suggests that in the beginning stages of industrialization, environmental conditions tend to deteriorate because the country is focused on economic development. This trend, however, is predicted to reverse once basic needs are met and nations reach middle-income levels per capita Gross Domestic Product (GDP) of US \$5000–\$8000.[7] According to Yandle et al.[7] and other scholars, countries will take more interest in cleaning up their environments once GDP reaches a certain level. This is the turning point when the line begins to fall (see Fig. 1). The curve differs according to the pollutant,[7] but there is no single environmental Kuznets Curve that fits all pollutants for all places and all times. The conclusion is that once some environmental degradation is accepted during industrialization, economic growth can be the force that helps undo previous environmental damage.

The economic perspective of this analysis concludes by stating that if economic growth is good for the environment, policies that stimulate growth (economic restructuring, trade liberalization, and price reform, for example) should be good for the environment. Overall, the conclusion is that once a country reaches a certain level, pressure will build up to protect the environment via policy measures and consumer behavior.[7]

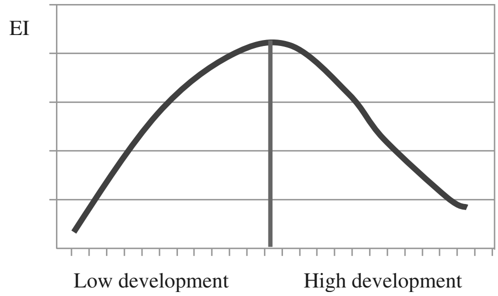


Fig. 1 Kuznet's curve for environmental impact (EI) as a function of development.

ENERGY PROJECTS MANAGEMENT

One of the most important aspects of an EM project is strong support from upper management. If this is not in place, the program will most certainly fail. The literature is full of companies that have illustrated their experiences with such programs, which are all total successes. Many companies find it difficult to pay for energy-efficient measures, even though they will save a significant amount of money in the long run. Below we summarize a set of opportunities to finance the implementation of energy savings projects, which are also discussed in other chapters in this energy encyclopedia.

In-house capital	
Utility incentive programs	
Debt financing/loans: (government loans)	The State of Florida Energy Loan Program (FELP)[6] requires that the interested party have 200 employees or less and have made a profit for the past two consecutive years
Commercial lending institutions	
Leases	Capital leases
	Operating leases
Performance contracts	Shared savings
	Performance contract for energy services (ESCO)
	Insurance guarantee of energy savings

THE ENERGY MANAGEMENT PROGRAM AND THE AUDIT PROCESS

No matter the type of facility, an energy audit program will always find savings through more efficient means of manufacturing and operations—the cost of the projects being the key point. To save energy, isolated projects are usually undertaken. However, such an endeavor should adhere to an energy program that must consider an energy assessment at the facility that will quantify energy and money savings to be obtained through a thorough engineering evaluation. To achieve these goals, this

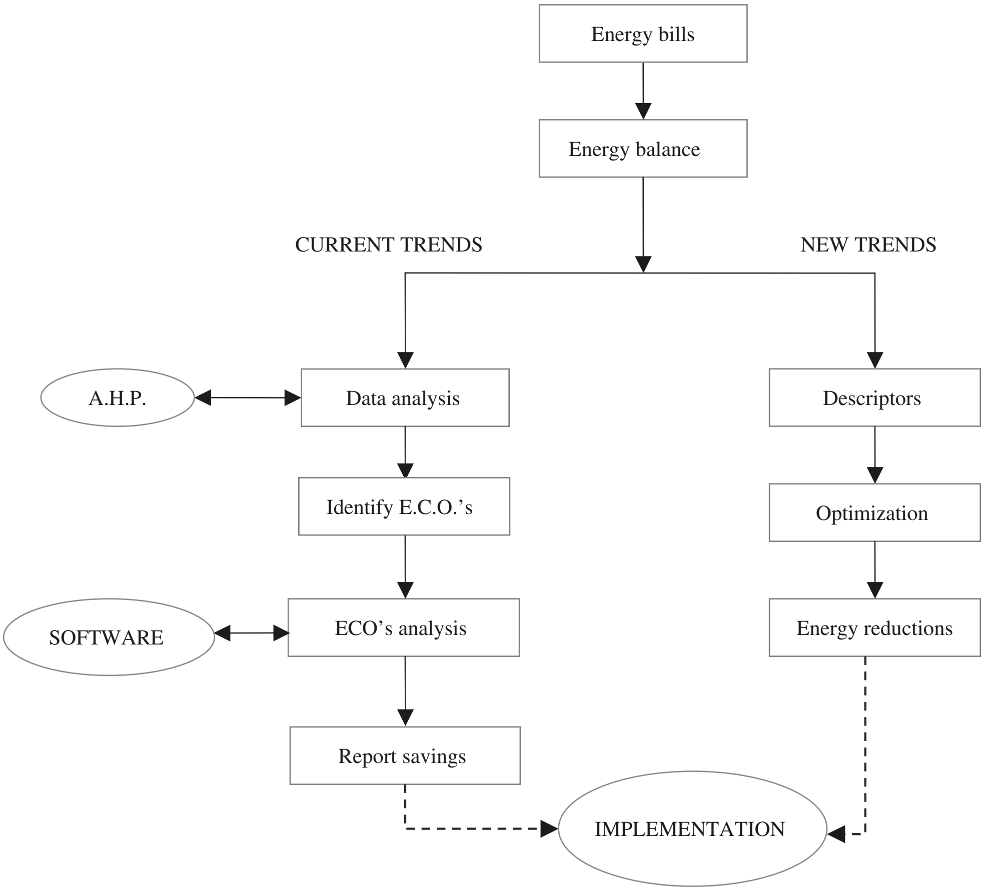


Fig. 2 The energy management (EM) program, including current and new trends. Notice that both encompass an EB.

chapter will show the importance of an energy balance (EB) in the EM program. The EM program scheme shown in Fig. 2 considers the analysis of current energy bills (to seek patterns and obtain real costs of demand and energy), an EB and, from there, the options for proceeding through current and/or new trends to pursue and implement energy savings projects. No matter the route followed, both consider the EB to be mandatory.

It must be pointed out that both trends pass through an EB and consider the implementation of the ECOs. It is an extra step to iterate again so as to improve until satisfaction levels are reached according to a pre-established energy conservation master plan.

Energy Balance: its main goal is to account for the electric energy consumed by each piece of equipment in a facility. This, in turn, allows for the performance of EM studies that track energy consumption for lighting, air compressors, air conditioning, motors, and any other piece of equipment. The energy consumed by each piece of equipment is then balanced against the facility's overall energy consumption (this allows the researcher to look for trends). The importance of an EB has been stressed in a set of published papers.[8–13] Recently, an EM program that does the balancing has been created in the form of software.[10–13] The software generates a report with the balance, graphs, and pie charts, and a set of ECOs that can be quickly evaluated. The importance of the EB is based on avoiding the not-uncommon situation of having a project that saves more energy than what the equipment currently requires. We shall return to the EB and its importance when we discuss some frequently used software.

Current Trends: today, no serious energy assessment can be completed without an EB that, through an Assisted Hierarchical Process (AHP), emphasizes the analysis of equipment efficiency and its loads. Then, as an inspection of the current process is performed, ECOs are identified and analyzed with the help of some software (to be discussed later). Finally, those ECOs of interest are reported.

New Trends: these are initiatives that come into place from an operations perspective so as to secure further savings. They are a function of the EB, and comprise a good procedure if the current trends ECOs have already been explored. Through a set of indicators, which we call descriptors,[14] we will discuss new trends that can, in turn, be used as benchmarks.

ENERGY SAVINGS: A CASE STUDY

We will consider Florida orange and grape industry—one of the largest manufacturing sectors in this state—energy usage[14] as a case study. The cost of manufacturing citrus juice depends on the cost of growing and picking the fruit from the groves, extracting fresh juice, and producing concentrate. We will show that a review of operations can help to improve procedures, and efficiency, and generate interesting savings. In this study, we analyze the general background of these facilities, their energy usage patterns, and the recommendations made to reduce energy consumption and production costs, as well (less manufacturing cost per unit). The citrus juice manufacturing processes performed in these facilities is succinctly illustrated in Fig. 3. Rather, we concentrate on the processes with a focus on energy usage requirements, although waste management and productivity issues should also be included.

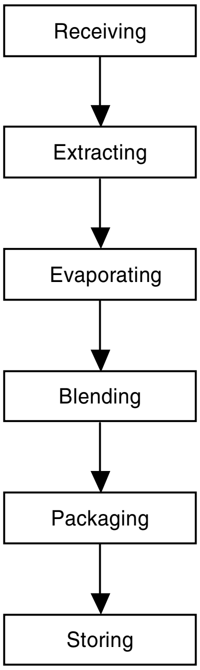


Fig. 3 Citrus juice manufacturing process.

Citrus Juice Production

In general, as well as after receiving, the fruit is washed with a soap solution while on the conveyors. The quality of grapefruits is tested by the U.S. Food and Drug Administration (FDA) through an on-site office. The FDA officers test the quality of fresh fruit coming into the facility. There is also some visual inspection, which is a crucial part of the process not only from the perspective of quality control, but also from the regulatory stand point.

In Table 1, we present the profile of four facilities with Standard Industrial Classification (SIC) code 2033,[15] sales figures, annual production, facility size, production cost, and number of employees. All four facilities participate in the State of Florida Electricity Tax Exemption program.[5] Also displayed are the energy savings (the number in parenthesis indicates the number of ECOs made during the assessment). The process is briefly pinpointed in Fig. 3. Finally, the percentage that these ECOs represent with respect to total electric energy consumption is also displayed. Other related citrus industries have been examined, too (SIC 2037 and 2087), and we refer the reader to the literature.[14]

Energy Consumption

The energy consumption of the facilities differs from one to the other, even if they manufacture the same product. Nevertheless, a closer look at their energy consumption reveals some trends. It also allows the relative merits of energy conservation and load management to be assessed. In this section, we will discuss these trends in terms of the energy used by type of equipment, and other features, like energy costs. An energy profile of the four facilities studied is shown in Table 1.

The Utility Company

While most of the large utilities offer good financial incentives to companies that install energy and demand saving equipment, midsize utilities offer little incentive for the adoption of new technology (like solar energy for water heating), and small utilities offer no incentives. It is

Table 1 Energy profile of four citrus companies

			Production area					
Plant	Sales ($/year)	Annual production	ft^2	\$/ft^2	EMPs	Taxes prog.	E. savings ($/year)	% of E savings
A	45,000,000	7,000,000 cases	142,283	316.27	175	Yes	86,741 (6)	10.83
B	50,000,000	25,000,000 lbs of Conc.	300,000	166.67	150	Yes	216,126 (5)	26.98
C	100,000,000	14,000 boxes fruit juice	645,712	154.87	150	Yes	124,460 (8)	15.54
D	140,000,000	3,400,000 gals of Conc.	100,341	1,395.24	140	Yes	189,709 (9)	23.68

Table 2 Energy profile prior to energy audits

	Electricity					Natural gas		Energy costs
Plant	KW	kWh	$/kW/month	$/kWh no demand	$/kWh with demand	MMBtu/year	$/MMBtu	per sales (%)
A	3,656	21,904,186	2.17	0.041	0.045	13,863	5.60	2.35
B	4,674	15,754,599	2.45	0.041	0.061	269,913	2.65	3.28
C	6,689	33,301,959	1.66	0.043	0.047	241,438	7.83	15.19
D	2,125	10,079,587	8.59	0.064	0.083	234,553	1.10	0.80
Average	4,286	20,260,083	3.72	0.047	0.059	189,942	4.295	5.40

indeed a very good practice to stay in close touch with the utility company and its available programs.

Energy Data

Table 2 shows energy usage and the cost of electricity and natural gas for the same four facilities. It becomes clear that, on average, the cost of electricity is very reasonable. However, there are still some exceptions due to small electric utilities, as shown by the rates of plant D. This difference sometimes cannot be changed by the facility, as it is geographically dependant. Would de-regulation help to reduce these values? At the end of 2002 and the beginning of 2003, the cost of natural gas increased about 300%, which resulted in a lagging cost increase in electricity. Now, a day's natural gas has practically the same cost as electricity in terms of dollars per MMBtu. Would a combined heat and power system[12] be a more attractive solution, particularly to serve end-of-the-line areas? This question will remain an issue.

The average values for demand and energy costs are shown in Table 2 under demand and energy costs. These costs are used in the economic analysis of the ECOs. These costs include taxes and are all based on a 12-month period, and, consequently, they differ from those shown on energy bills because ours include all other charges as applicable taxes, fuel cost adjustments, etc.

ENERGY BALANCE AND ENERGY COST DISTRIBUTION

As a regular practice, and to identify electric energy usage by equipment type, we developed EB software.[8–13] Fig. 4 shows the average equipment energy usage distribution from where it becomes clear that chillers, motors, air conditioners, and ammonia compressors are the main electric energy consumers. These facilities are much larger electric energy consumers than related industries (SIC 2037 and 2087, for example) by approximately a factor of five. As noted above, the inconsistent cost of natural gas makes it rather difficult to attempt to procure averages. Therefore, we concentrate our analysis on electrical energy consumption only. In this type of industry, natural gas is mostly used by kilns for the drying process.

CURRENT TRENDS: ECOs

As the energy assessments were performed, many of the recommendations summarized in Table 3 were made for each plant. These facilities have energy usage seasonal

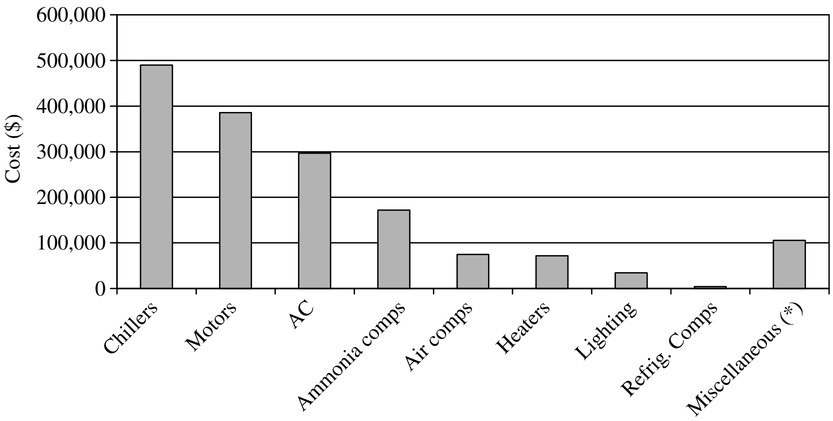


Fig. 4 Average equipment energy usage. (*Miscellaneous considers all minor equipment as computers, copiers, printers, desk-lamps, etc.).

Indu–Inter

Table 3 Energy conservation opportunities

Assessment recommendations	
Install high efficiency lighting	Repair steam leaks
Install combined heat and power system	Recover waste/exhaust heat
Install new refrigeration system	Insulate facility roof
Install high efficiency motors	Insulate roof of tank farms
Replace V belts with cogged V belts	Turn off lights and AC
Turn off unused equipment	
Install occupancy sensors	
Repair compressed air leaks	Remove unnecessary lights
Provide cooler air for air compressor	Schedule freezers lights use
Reduce compressed air pressure	Turn off outside lights
Reduce lighting in selected areas	Install skylights and sensors
Retrofit gas fired system	Implement a lights management system
Adjust boiler air/fuel ratio	Change feed mill schedule
Increase frequency of boiler tune-up	Install a light-bulbs recycling unit, etc.

fluctuations due to the nature of their business. Additional savings recommendations, which are not listed here, were made in the areas of waste handling and productivity enhancement.

Data Analysis

Based on the data shown, we are now in a position to estimate potential savings for the type of facilities considered, with the caveat that they are suggestions that show trends (and that they represent a fraction of this Florida industry). To show potential savings, we consider the average of energy savings and the fraction that corresponds to the energy usage for each type of equipment listed.

Consequently, we can write the fraction of savings $F_{i,j}$ as follows:

$$F_{i,j} \;=\; \langle S_j \rangle \left[C_j / \sum_{j=1} T_j \right]$$

where $F_{i,j}$ = Counters: i is for kWh = 1; kW = 2; cos t ($/year) = 3, j is for the type of piece of equipment. It runs from 1 to 11 as of the sequence shown in Table 4 (first column). $\langle S_i \rangle$ = Average of Savings (kW, kWh, cos t), C_j = Current Consumption (kW, kWh, cos t), T_j = Total Current Distribution according to piece of equipment.

In this fashion, we see how the energy is being distributed, and hence, we obtain proposed energy, demand, and associated cost savings that we can now directly compare with current usage. We perform our analysis by type of energy consumers that are typically present in most manufacturing facilities, listed in Table 4. Also listed are the average proposed energy (kWh), demand (kW), cost ($/year) savings (used to compute the associated cost savings), associated implementation cost, the expected simple payback period (SPP), and the return on investment (ROI), all of which are a consequence of the energy savings and demand reduction (DR).[14]

Table 4 Savings from the energy assessments

	Savings					
	kWh	kW	$/year	Implementation cost ($)	SPP years	ROI %/year
Lighting	111,433	29	6,546	14,274	2.2	46
Motors	1,136,974	52	56,087	30,181	0.5	186
Air cond.	1,210,020		66,914	75,000	1.1	89
Air compressors	284,033		15,707	1,050	0.1	1,496
Ammonia compressors	541,672		29,954	45,000	1.5	67
Hydraulic comps.						
Refrig. Comps.	17,242		953	2,500	2.6	38
Chillers	2,108,067		116,576	40,000	0.3	291
Heaters	77,582		4,290	31,741	7.4	14
Dehumidifiers	1,557,563	555	98,459	15,000	0.2	656
Miscellaneous	416,040		23,007	16,000	0.7	144
Total savings	7,460,625	636	418,494	270,745	0.6	155

NEW TRENDS: DESCRIPTORS FOR ENERGY SAVINGS

In this section, we analyze how electric energy is used by different pieces of equipment in the facilities. Obviously, some consume more energy than others and, as a consequence, are more relevant for an energy conservation study, and to describe energy consumption in the plant. Accordingly, we recognize this by first classifying the energy consumers by the kind of equipment, as we did in Table 4 (1=Lighting; 2=motors, 3=AC, etc.), and shown in Table 5. Secondly, we assign them a specific weight, whose value will serve to describe the electric energy consumption; we will call them electric energy usage descriptors.[14]

We compute these descriptors ($f_{i,k}$) by taking the average consumption value of a given type of equipment $\langle L_{i,k} \rangle$ and dividing it by the sum of all. This is:

$$f_{i,k} = \langle L_{i,k} \rangle / \sum_{i=1}^{11} L_{i,k}$$

Notice that the counter i runs for all equipment considered and shown in Table 5. The counters are for demand ($k=1$), energy ($k=2$), and cost savings ($k=3$). These descriptors must satisfy the following necessary condition:

$$\sum_{i=1} f_{i,k} = 1$$

The calculated values of computed descriptors for demand, energy, and costs are shown in Table 5. The highlighted values again indicate the major contributors that better describe the energy usage in the plants studied. Motors, air conditioning, chillers, and compressors are the big consumers. On average, this equipment represents 87, 73, and 82% of all the electric demand, energy, and associated manufacturing costs.

These descriptors multiplied by the demand, energy, or cost associated with the equipment (lights, motors, chillers, etc.), account for the total electric energy use. However, and as discussed above, there are some particular pieces of equipment that account for most of the descriptor (demand, energy, and cost savings).

For this case, the demand, energy, and cost descriptors are (from Table 5):

Demand ($k=1$):

$$f_{i,k} = f_{i,1} = f_{2,1} + f_{3,1} + f_{5,1} + f_{8,1}$$
$$= 0.255 + 0.209 + 0.107 + 0.298 = 0.869$$

Energy ($k=2$):

$$f_{i,k} = f_{i,2} = f_{2,2} + f_{3,2} + f_{8,2}$$
$$= 0.185 + 0.197 + 0.344 = 0.726$$

Cost ($k=3$):

$$f_{i,k} = f_{i,3} = f_{2,3} + f_{3,3} + f_{5,3} + f_{8,1}$$
$$= 0.236 + 0.182 + 0.105 + 0.300 = 0.823$$

These descriptors describe the parameter of interest (demand, energy, and cost savings). The advantage is that we focus only on a few types of equipment. Notice that here only those components bigger or equal than 0.100 are considered.

Table 5 Descriptors distribution

	kW	kWh	Cost ($)
(1) Lighting	0.017	0.018	0.021
(2) Motors	0.255	0.185	0.236
(3) Air Cond.	0.209	0.197	0.182
(4) Air Comps.	0.045	0.046	0.046
(5) Ammonia compressors	0.107	0.088	0.105
(6) Hydraulic compressors			
(7) Refrigeration compressors	0.002	0.003	0.002
(8) Chillers	0.298	0.344	0.300
(9) Heaters	0.048	0.050	0.044
(10) Dehumidifiers			
(11) Miscellaneous	0.017	0.068	0.064
Total	1.000	1.000	1.000

SOME AVAILABLE SOFTWARE

In this section, we provide a very simple, quick review of relevant software, as provided by the U.S. DOE Best Practice program. The different software are useful in helping the user make energy conservation decisions in the areas of motors, pumps, compressed air, steam, etc. (these program packages are free to download from the Web).[16] They are very versatile and easy to use. Data entries like location, utility details (energy costs), the system itself, efficiency, maintenance, life cycling, etc. are required. They are as follows.

Motor Master (MM+4.0)

Motor Master is a package designed to help utility auditors and industrial energy coordinators support motor improvement planning by identifying the most efficient action for a given repair or motor purchase action.

It contains an amazing amount of information for over 10,000 inventory motors from 15 different motor manufacturers.

Air Master +1.0.9

Air Master has a large suite of compressed air offerings that allow the user to explore possibilities for improving the compressed air system performance and operation through energy efficiency measures.

PSAT: Pumping Assessment Tool

Being very thorough software, PSAT will compute, according to input data, the optimal energy-efficient pump motor for the user, given a user optimization rating (input).

Steam System Assessment Tool 1.0.0

Through the analysis of operational steam-systems data, the auditor can examine possible savings in energy, cost, and emissions.

Steam System Scoping Tool 1.0d

Designed to be used by steam systems operations personnel for large industrial plants, this package offers good operational benchmarking, and it comes in the form of an Excel spreadsheet that allows the user to establish a record of what is currently being done in the steam system so as to show potential opportunities for improvement, evaluate current operational procedures, and compare steam systems with those of other facilities.

3E Plus

3E Plus is a very useful industrial insulation evaluation tool for exploring the best economic alternatives.

Process Heating Assessment and Survey Tool (PHAST 1.1.1)

This is a very valuable tool that helps to improve thermal efficiency on process heating equipment. For example, it allows for the investigation of cascading possibilities for process heat.

Industrial Energy Management Software

In a previous work, we described the new interactive EB software.[8–13] The package offers an easy way to account for energy bills, working equipment, and their energy operating characteristics: lighting, air conditioning, refrigeration compressors, air compressors, motors, and others. The electric energy consumed by all pieces of equipment is balanced against the energy bills. The corresponding energy costs, graphs, pie charts, and free format reports are automatically generated and show paths and trends of energy consumption. A "Recommendations" section allows for a quick, but beyond-the-back-of-the-envelope type of calculation, initial analysis of potential energy improvement projects associated with current equipment, and/or for the installation of new ones. After entering electric bills and equipment data, it performs an EB for the user.[10] This is a list of all the equipment that utilizes electricity as well as consumption operational patterns. The total is finally compared with the actual electric energy bills (usually from the previous 12 months), requiring a deviation of less than or equal to 2%.

To deal with other important pieces of equipment, we refer the reader to other chapters in this encyclopedia and other references, as well.[17–22]

DISCUSSION

From the proposed 7,460,625 kWh/year savings for these industries, we realize that not only is the bottom line affected, but the environment is, as well. Taking an optimistic perspective, let us assume that this electricity was generated with an efficiency of 40%, and an implementation rate of the same fraction, meaning that about 25,500 MMBtu/year of fuel are to be saved. To put this in the proper context, this represents, on an annual basis, unburned fuel equivalent to 255,000 therms of natural gas, or 5000 barrels of crude oil, or 1020 ton of coal. This is just the good news; the bad news is that 1.5 times thisamount is sent to the atmosphere, with all of the associated pollutants. This phenomenon is happening constantly all over the world in a kind of export fashion of the nontraditional type. Some utilities do have the appropriate required filters (such as scrubbers), but they are not enough.

NEW AND ALTERNATIVE SOURCES OF ENERGY

Today, the high costs of energy call for new sources of energy that pass through renewables and alternative fuels that are not generally considered, but that are sometimes process-dependent. Some of these are cars and trucks tires, which burn in an oven that handles them in a fork display; industrial oils, which can be further treated to eliminate their excess of water; and chemicals like sulfur, which, when burned, generates power and, as a by-product, provides sulfuric acid with many applications (manufacture of fertilizers, for example).

Another use of energy is symbiotic—this is a partnership between different types of industries in which one

Table 6 Compare demand, energy and cost descriptors distribution for related SIC codes 2037 and 2087

		2037			2087	
	kW	kWh	Cost ($)	kW	kWh	Cost ($)
(1) Lighting	0.046	0.067	0.071	0.030	0.051	0.027
(2) Motors	0.511	0.35	0.428	0.383	0.344	0.359
(3) Air Cond.	0.021	0.011	0.015	0.017	0.008	0.010
(4) Air Comps.	0.026	0.045	0.050	0.010	0.007	0.008
(5) Ammonia Compressors	0.339	0.425	0.310	0.203	0.232	0.245
(6) Hydraulic Compressors				0.250	0.260	0.250
(7) Refrigerator Compressors						
(8) Chillers	0.042	0.063	0.091	0.031	0.015	0.015
(9) Heaters						
(10) Dehumidifers	0.015	0.033	0.036			
(11) Miscellaneous				0.076	0.083	0.085
TOTAL	1.000	1.000	1.000	1.000	1.000	1.000

generates waste that is fuel for the other. An example is biogas, where the manure (waste) generated on an animal farm can become the fuel source (biogas) of any facility that can use gas in its process. In this case, symbiosis has the added value of solving a health problem that results from piling-up excreta. These symbiotic possibilities should be explored thoroughly by energy managers.

CONCLUSIONS

We have presented a review of the current energy usage patterns of a number of Florida juice (orange and grapefruit) manufacturing facilities at which the University of Florida IAC has performed energy audits. Their current energy (heat and power) usage patterns, associated costs, energy indicators, and recommended energy savings measures have been discussed.

Through an analysis of the available data, we have presented a consistent set of measures that can be applied to reduce the production cost of citrus juice through the establishment of energy-saving measures or simple energy cost reduction. Moreover, and for the SICs considered, we have identified the areas that should be prioritized through a descriptors approach.

In our study, we have proposed the use of descriptors, whose advantages include narrowing down the areas in which an EM program should concentrate. In other words, we have made a Pareto identification of energy consumption areas to which priority should be given. Moreover, we believe that through this process we are also providing a tool for benchmarking the considered SIC codes. In essence, we foresee that facilities with SIC codes, like those considered here, can easily identify savings possibilities and, by comparison, benchmark their operations. Tables 5 and 6 provides the necessary information for this task, for different but related Standard Industrial Classification (SIC) codes (2033, 2037 and 2087, respectively).

Heat recovery is an attractive idea, as the cost per MMBtu coming from electricity is currently approximately (average) 2.5 times higher than the one from natural gas. As discussed at the very beginning of this entry, the cost of natural gas has gone up, but we believe that this fact should encourage heat recovery opportunities, as recommended to practically all the plants considered here. Finally, energy conservation opportunities do provide pollution reduction in a somewhat direct mode. Indeed, electricity is an on-time commodity, hence each time a plant reduces its electric energy consumption, the corresponding utility company burns less fuel with the consequent pollution reduction. Hence, this work, like others in EM, has a dual purpose—energy savings and environment conservation.

ACKNOWLEDGMENTS

The author would like to acknowledge helpful discussions with Mrs. Alejandra Bórquez and Professor Barney Capehart, the editor of this great encyclopedia of energy.

REFERENCES

1. Allen, D.T.; Shonnard, D.R. *Green Engineering: Environmentally Conscious Design of Chemical Processes*; Prentice Hall, Inc.: New Jersey, NJ, 2002.
2. 3M Company website: www.mmm.com (accessed on November 12, 2006). Allen, P.J. *How Disney Saves Energy and Operating Costs*; HPAC Magazine , 2005.
3. The United States Department of Energy: www.doe.gov (accessed on November 12, 2006).

4. The Federal Energy Management Program (FEMP): www.eren.doe.gov/femp Browse around this website and explore the possibilities; The University of Florida Industrial Assessment Center website: www.ise.ufl.edu/iac (accessed on November 12, 2006).
5. Florida Department of Revenue, P.O. Box 7443, Tallahassee, Florida 32314-7443.
6. Porter, M.E. America's green strategy. Sci. Am. **1991**, *264*, 168; Porter, M.E.; Van der Linde, C. Toward a new conception of the environment–competitiveness relationship. J. Econ. Perspect. **1995**, *9* (4), 97–118.
7. Esty, D.C. Bridging the trade-environment divide. J. Econ. Perspect. **2001**, *15* (3), 113–130; (www.sice.oas.org/geograph); Yandle, B.; Vijayaraghavan, M.; Bhattarai, M. *Environmental Kuznets Curve: A Primer*, PERC Research Study 02-1, 2002 (www.perc-org); Shafik, N.; Bandyopadhyay, S. *Economic Growth and Environmental Quality: Time Series and Cross-Country Evidence*, World Bank Policy Research Working Paper WPS904, 1992; Grossman, G.M.; Krueger, A.B. Economic growth and the Environmen. Quart. J. Econ. **1995**, *110*, 353–377; Webber, J.R.; Allen, D.O. *The Environmental Kuznet's Curve: Mess or Meaning?*, 2004 (http://carecon.org.uk); Reynolds, M.E. *The Business Implications of the Environmental Chapter of the US-Chile Free Trade Agreement*, Masters Degree Thesis in Latin American Studies at the University of Florida. U.S.A. 2006.
8. Pawlik, K-D.E.; Capehart, L.C.; Capehart, B.L. *Analyzing Energy Use: A Balancing Act*; Strategic Planning for Energy and the Environment: Lilburn, GA, 2001; (Fall).
9. Pawlik, K-D.E.; Capehart, L.C.; Capehart, B.L. *Energy Balancing—How to Use the Energy Balance Data You Have Collected to Make Financial Decisions*, Strategic Planning for Energy and the Environment: Lilburn, GA, Winter 2001–02, Vol. 21, No. 3.
10. Cardenas-Lailhacar, C.; Krishnamurthy, S.; Schaub, D.; Kirmse, D. *An Interactive Energy Balance: A Case Study*, Proceedings of the 2001 ACEEE Summer Study on Energy Efficiency in Industry. Vol. 1. Tarrytown, NY, July 2001.
11. Cardenas-Lailhacar, C.; Venkat, S.; Schaub, D.; Capehart, B.L.; Capehart, L.C.; Paulik, K.-D.E. *An Energy Savings Case Study Using the Interactive Energy Balance*, Strategic Planning for Energy and the Environment: Lilburn, GA, Vol. 25, No. 1, Summer 2005.
12. Cardenas-Lailhacar, C.; Serrano, H. *A Cogeneration Case Study Using the Interactive Energy Balance Program*, World Energy Engineering Congress WEEC-2005, Austin, TX, September 2005.
13. Cardenas-Lailhacar, C.; Domanico, L.; Latorre, L. *A New power Factor Correction Setup Using an Industrial Energy Management Program*, World Energy Engineering Congress WEEC-2006, Washington, DC, September 2006.
14. Cardenas-Lailhacar, C.; Corrochano, E.; Lescano, J.; Schaub, D.; Serrano, H. *Energy Savings in the Citrus Industry: Proposed Solutions*, 2003 Summer Study on Energy Efficiency (ACEEE) International Conference, New York, July 29–August 1, 2003.
15. *2006 Directory of Florida Industries*; Harris Infosource: a D & B Company, 2057 E. Aurora Road, Twinsburg, OH. 44087-1999. ISBN 1-55600-268-8.
16. U.S. Department of Energy, Office of Industrial Technologies. Free software that can be downloaded from: www.oit.doe.gov/bestpractices/ (accessed on November 12, 2006). Also see: Cardenas-Lailhacar, C. Chapter: *Interactive Computer Tools for Energy Auditing*, Proceedings Business Energy Solutions Expo. Orlando, USA. November 2002.
17. This website gives the tonnage of the unit, provided that the serial and model numbers are known (i.e., obtained from the name-plate): www.ARI.org (accessed on November 12, 2006)
18. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 5th Ed.; The Fairmont Press Inc.: Lilburn, GA, 2005; Turner, W.C. *Energy Management Handbook*, 4th Ed.; The Fairmont Press, Inc.: Lilburn, GA, 2005; Turner, W.C. *Energy Management Handbook*, 3rd Ed.; The Fairmont Press, Inc.: Lilburn, GA, 1997; Turner, W.C. *Energy Management Handbook*, 3rd Ed.; The Fairmont Press, Inc.: Lilburn, GA, 1993.
19. Illuminating Society of North America. *Lighting Handbook, Reference & Application*, 8th Ed.; IESNA: New York, NY, 1993.
20. 1999 State Fact Book, Office of Industrial Technologies, Office of Efficiency and Renewable Energy. U.S. Department of Energy. Also see: EPRI Journal, 1992.
21. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 3rd Ed.; The Fairmont Press, Inc.: Lilburn, GA, 2000. Turner, W.C. *Energy Management Handbook*, 3rd Ed.; The Fairmont Press, Inc.: Lilburn, GA, 1997. Dr. Barney Capehart's website: www.ise.ufl.edu/capehart/ (accessed on November 12, 2006).
22. The Association of Energy Engineers (AEE) Web site: www.aeecenter.org/index.html (accessed on November 12, 2006).

Industrial Motor System Optimization Projects in the U.S.☆

Robert Bruce Lung
Resource Dynamics Corporation, Vienna, Virginia, U.S.A.
Aimee McKane
Lawrence Berkeley National Laboratory, Washington, D.C., U.S.A.
Mitchell Olszewski
Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.

Abstract
This study examines data from 41 industrial motor system optimization projects implemented between 1995 and 2001 that were developed into Department of Energy (DOE) case studies to determine the effect of the energy savings brought about by such projects. The study calculates aggregate energy savings, the net present value (NPV) of project savings, and the internal rates of return of project cash flow (IRR) for each project, as well as for all projects aggregated together and for projects having large capital expenditures vs. those in which the systems were optimized primarily with engineering changes. Finally, the study makes some rough estimates of possible energy savings throughout U.S. industry.

For this study, projects are considered financially worthwhile if the NPV of the project savings over a 10-year project life is greater than the project cost. Because the simple payback criterion is the norm in the DOE case studies, it is used as a complementary measure of success.

The initial results suggest that the NPV for most of the projects' savings in the sample are positive with 10- and 15-year project lives, and that many are even positive with 5-year project lives. For projects involving large capital purchases, a project life of 10 years is required for a positive NPV, while for projects that primarily involve engineering changes, a positive NPV was achieved with a 5-year project life. This suggests that motor system optimization projects do not necessarily require large capital expenditures to achieve substantial energy savings.

INTRODUCTION

Since the mid-1990s, manufacturing plants in the United States have undertaken measures to reduce energy consumption. One of the primary types of these measures has been to increase the efficiency of their motor systems. These efforts have been encouraged by the U.S. DOE Office of Energy Efficiency and Renewable Energy (EERE) through its Best Practices program. The Best Practices program includes the publication of case studies and similar documents (approximately 150 since 1995) on industrial energy efficiency projects to raise awareness within industry of the benefits of such projects.

This study explores the effect of such projects by examining data from 41 separate projects that were developed into DOE case studies. Through this review, it is expected that a better understanding of the impact of industrial motor system improvement projects will be achieved.

This article is organized in the following manner. The first part defines and discusses the methodology of the study. Next, the analytical tools employed are discussed, along with the data and the selection criteria. In the next section, the results are presented, discussed, and interpreted. Finally, some concluding remarks are presented.

Methodology

The main objective of this study is to estimate the impact of motor system optimization on industrial competitiveness and to provide insights into the mechanisms by which such benefits are delivered to industry. Specifically, it is expected that the study will (1) provide planning-relevant information on the nature and magnitude of the financial impacts from industrial motor system optimization, and (2) convey to the policy process, the rates of return to U.S. industry from expenditures on such efforts. The data from the 41 projects are aggregated to provide a collective estimate of their impact and extrapolated to estimate potential industry-wide energy savings, which will allow the results of this assessment to feed back into DOE's strategic planning process.

Empirical Measures

The primary quantitative measures employed in this study are NPV, IRR, and simple payback. The NPV is used

☆ This entry originally appeared as "Industrial Motor System Optimization Projects in the U.S." in *Energy Engineering*, Vol. 102, No. 1, 2005. Reprinted with permission from AEE/Fairmont Press.

Keywords: Energy efficiency; Industrial; Productivity; Motor systems.

because it is an absolute measure of the value or expected value of an investment in constant dollar terms. It therefore portrays the annual energy savings from the projects as interest payments on a bond, and can be used to compare the performance of expenditures on motor system improvement projects to the performance of other potential uses of financial capital. If the 10-year NPV of project savings exceeds the project cost, the project will be considered successful.

The IRR of project cash flow is the compound rate of return paid by an investment. In this study, the project cost is the investment and the IRR of project cash flow gives each project's total return on its project cost. By providing a rate of return measure, the IRR of project cash flow allows comparison with a designated opportunity cost of the funds being considered for the projects. For this study, an IRR of project cash flow that is greater than 50% will constitute a rate of return that makes a project successful. For the aggregate data, the IRR will be labeled the CRR of project cash flow (collective rate of return).

The simple payback is a complementary measure that serves as a modified benefit-cost ratio. According to the U.S. Industrial Motor Systems Market Opportunities Assessment (MSMOA), many firms use only the simple payback as a decision tool (MSMOA, P. 54), and DOE uses it in all of its case studies, making it relevant for this study. Although the MSMOA applies a 3 year simple payback to energy efficiency measures, some of the firms whose projects are used in the study report using a financial hurdle rate of 2 years or less on the simple payback to obtain project approval. Therefore, projects having a simple payback of 2 years or less will be considered successful in this study.

In addition to energy savings, many of the projects in the study have led to important non-energy benefits such as maintenance savings, increased or improved production, and lower emissions. These additional benefits are in effect positive externalities of motor system improvements. With these methods, this study provides both qualitative assessments and quantitative estimates of the net financial benefits resulting from industrial motor system optimization.

Data

The data used in the study come from 41 separate motor system optimization projects that were implemented in manufacturing plants in the United States between 1995 and 2001 and cover seven different motor systems across 17 industries. The 41 projects were selected because they all provided the following data: energy savings in U.S. dollars, energy savings and consumption in kilowatt-hours (kWh), project cost, and power cost ($/kWh). These projects are not a random sample; nor do they present the largest or smallest energy savings among motor system projects. However, in many cases their manner of implementation is consistent with best industry practices for motor system optimization.

Along with the above-mentioned attributes, additional assumptions in the study include a discount rate of 5%, and a project life of 10 years. The 5% discount rate is used because that rate approximates the average yield on the 10-year U.S. Treasury bond between 1995 and 2001. The 10-year project life is the typical life span reported by various consultants, manufacturers, and personnel involved in many of the study's projects. Because a 10-year project life may not be universal, results for 5- and 15-year project lives are also presented to see how the results differ with time periods that are 50% shorter and 50% longer than the selected baseline.

Results

The aggregate results show positive values for each of the measurement tools applied to the data. The aggregate project costs for all 41 projects total $16.8 million, and the aggregate savings sum is $7.4 million and 106 million kWh. Using these figures, the NPV of project savings for the baseline analysis is $39,572,000, which exceeds the aggregate project cost. The CRR of project cash flow in this analysis is positive at 41% and the simple payback is 2.27 years. These results are summarized in Table 1. Fig. 1 shows energy consumption before and after the projects. The initial aggregate results are skewed by two projects that were implemented with the intent of improving process reliability rather than saving energy. For these two projects, the costs vastly exceed the energy savings, and they are viewed as outlyers. When these two projects are taken out of the sample, the NPV of project savings

Table 1 Aggregate data

Aggregate data	$N=41$	Without outlyers ($N=39$)
Total costs	$16,772,740	$15,072,740
Total savings	$7,401,700	$7,303,700
Median cost	$240,000	$188,000
Median saving	$115,000	$108,000
Kilowatt-hour	106,483,517	104,544,517
Simple payback	2.27	2.06
Collective rate of return	41%	46%
Present value	$56,344,762	$55,598,746
Net present value 5-year	$13,913,292	$15,207,003
Net present value 10-year	$39,572,022	$40,526,006
Net present value 15-year	$61,975,378	$58,200,783

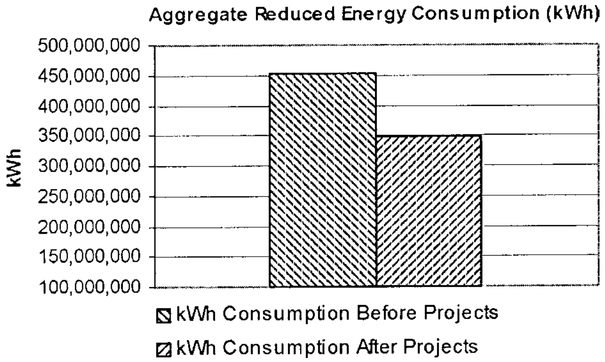


Fig. 1 Aggregate energy consumption before and after projects.

becomes $40,526,000, the CRR of project cash flow is 46%, and the simple payback is 2.07 years. Because of the outlyers, median rather than average values of project costs and savings are provided in Table 1. Using the aggregate data, the NPV and CRR are also calculated for 5- and 15-year project lives. For the 5-year project life, the NPV figure is positive but smaller than the total project cost. The CRR figure reduces to 32% and the simple payback stays the same. In the 15-year project life, the NPV greatly exceeds total project cost and the CRR of project cash flow is 43%.

Once the initial aggregate results were calculated, the data were divided according to two main sets of features. The two main sets of features were plants belonging to the industries of the future (IOF) or non-IOF companies, and projects involving much capital spending vs. others that depended primarily on re-engineering.

Projects at IOF vs. Non-IOF Industrial Facilities

Of the 41 projects in the study, 21 occurred in plants of companies that fall under DOE's IOF (mining, forest/paper products, steel, glass, metal casting, agriculture, petroleum, chemicals and aluminum.) designation, while 20 were implemented in facilities of non-IOF firms. Industries of the future companies are considered to have the most energy intensive industrial processes, and it was expected that motor system optimization projects in such facilities would have the greatest energy savings and highest returns. The results for the projects at IOF companies bear this expectation out and are displayed in Table 2, which shows that the NPV of project savings and IRR of project cash flow for the IOF facilities' projects are greater than those of the non-IOF firms. The before and after energy consumption patterns are shown in Fig. 2.

Table 2 Industries of the future (IOF) vs. non-IOF energy consumption

	IOF companies $N=21$	Non-IOF companies $N=20$
Total costs	$7,266,740	$9,506,000
Total savings	$4,182,700	$3,219,000
Median cost	$240,000	$219,000
Median saving	$120,000	$99,000
Kilowatt-hour	66,993,517	39,490,000
Simple payback	1.74	2.95
Internal rates of return	55%	31%
Present value	$31,840,420	$24,504,342
Net present value 5-year	$10,073,934	$3,839,358
Net present value 10-year	$24,573,680	$14,998,342
Net present value 15-year	$37,233,815	$24,741,564

Because the NPV and IRR figures of IOF facilities' projects are greater than those of non-IOF facilities, this suggests that motor system improvement projects in such facilities are desirable. The non-IOF facilities' lower NPV, IRR, and simple payback is partly because many of their projects incurred higher equipment costs than did the IOF

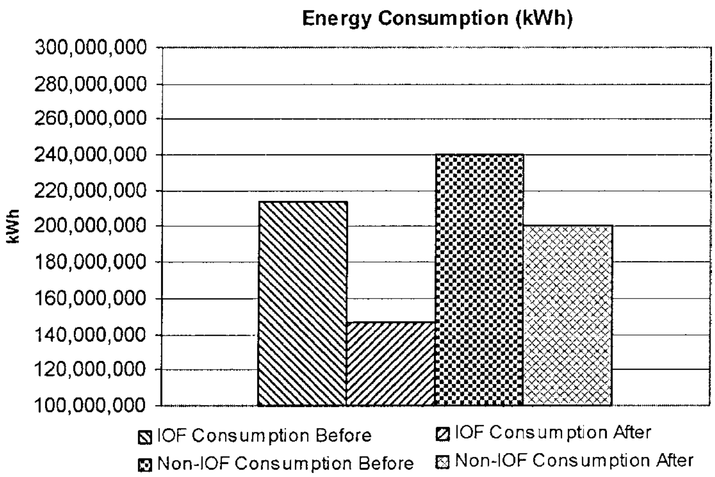


Fig. 2 Energy consumption before and after projects for industries of the future (IOF) and non-IOF companies energy consumption (kilowatt-hours).

Indu–Inter

Table 3 Capital spending vs. re-engineering

	Capital spending $N=35$	**Re-engineering** $N=6$
Total costs	$16,533,740	$239,000
Total savings	$6,999,700	$402,000
Median cost	$264,000	$35,000
Median saving	$115,000	$44,500
Kilowatt-hour	102,523,517	3,960,000
Simple payback	2.4	0.6
Internal rates of return	40%	164%
Present value	$53,284,574	$3,060,188
Net present value 5-year	$12,485,677	$1,427,615
Net present value 10-year	$36,750,834	$2,821,188
Net present value 15-year	$57,937,423	$4,037,956

facilities' projects. This higher cost is reflected by the fact that the weighted average of the non-IOF facilities' project costs is over 25% greater than that of the IOF facilities' projects.

Capital Spending vs. Re-engineering

The other salient feature about many of the projects in the study was the degree to which equipment replacement and spending on capital equipment was significant in many projects. Spending on capital equipment was equal to or greater than 70% of the total project cost in 31 of the study's 41 projects. Furthermore, the main features in 35 projects included the replacement or addition of OEM (original equipment manufacturer) devices such as pumps, fans, compressors, VSDs (variable speed drives), and control systems.

The projects were divided into two categories: capital spending and re-engineering. Projects that fell into the capital spending category were those for which most of the project cost was on new equipment or that were characterized by equipment replacement. The re-engineering category included projects in which capital equipment purchases were small relative to the projects' costs or in which new equipment did not significantly contribute to the projects' results. Of the 41 projects in the study, 35 were classified as capital spending and six as re-engineering.

The analysis for the two categories of projects produced vastly different results, which are shown in Table 3. Fig. 3 shows both groups' energy consumption before and after the projects. While the NPV of capital spending projects' savings makes them successful according to the study's criteria, their IRR of project cash flow and simple payback are below the study's criteria for success. This category contains the two outlyer projects, and when they are subtracted, the IRR of project cash flow and simple payback for the category remain below those criteria at 43% and 2.18 years. By contrast, the projects in the re-engineering category are successful according to all three metrics. Also, the NPV of project savings for the re-engineering projects is greater than the total project costs under the 5-year scenario.

While the data from the re-engineering projects in this study show that higher rates of return were obtained through reconfiguration of or adjustments to existing motor systems, this does not indicate that equipment replacement is not justified. In many of the capital spending projects, the new OEM equipment was smaller than the devices it replaced. Also, the new equipment was more technologically advanced or contributed to more efficient operation of the motor systems involved, which

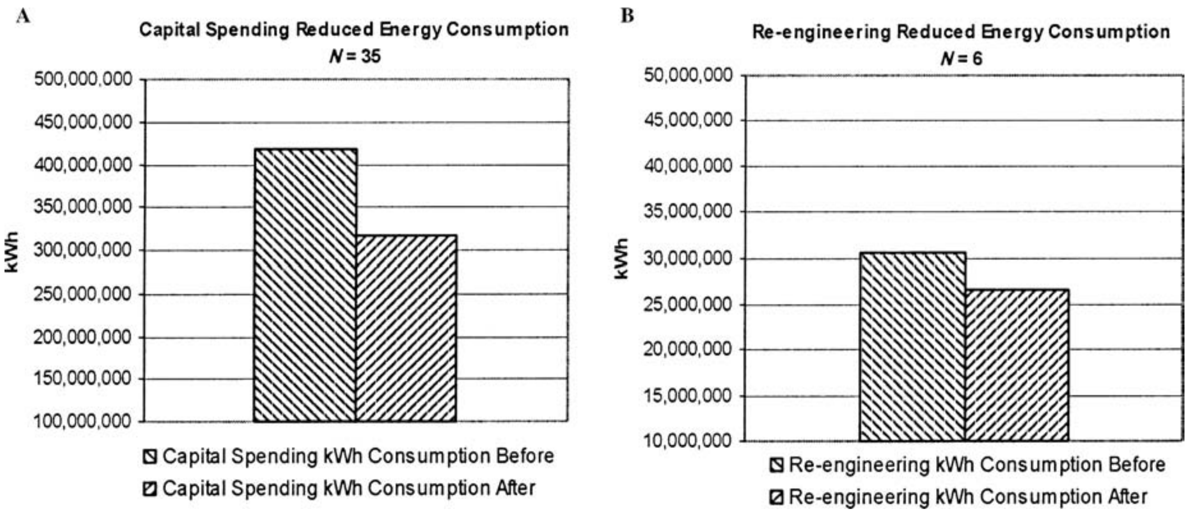


Fig. 3 (A) Energy consumption before and after projects; (B) Capital spending and re-engineering.

obviated the need for previously anticipated equipment purchases. However, the re-engineering projects' results do suggest that large capital equipment expenditures are not a necessary condition for energy efficiency.

The results in Table 3 display the NPV of project savings for 5- and 15-year project lives to show how the rates of return differ with time periods that are 50% shorter and 50% longer than the 10-year baseline project life. However, these time frames may not be appropriate. The 5-year scenario is very short for the capital spending projects, and the 15-year project life may be unrealistic for the projects that fall under the re-engineering category because it is unlikely that the results will persist for that length of time. Therefore, the more appropriate time-frames to view the NPV of project savings for the capital spending projects are the 10- and 15-year ones, while for the re-engineering projects, the 5- and 10-year NPVs would be more realistic and reflective of the findings obtained in the study.

Individual Project Results

All 41 projects were individually analyzed under 5-, 10-, and 15-year project lives. Table 4 shows the results along with the percentages that those results represent.

As Table 4 shows, the NPV of project savings exceeds the project costs in a majority of the 41. The IRR of project cash flow is greater than 50% in each project life scenario. This is also true for the projects that fell under the capital spending label, and particularly so for the re-engineering projects.

Indirect Impacts

While the vast majority of the projects provided measurable energy savings, some yielded unanticipated benefits that can be seen as positive externalities. Typically, many additional improvements or savings were generated in one or more of the following areas: productivity (production and/or product quality), maintenance, emissions, ancillary products, plant safety, and avoided equipment purchases. In addition, some firms received rebates or incentive payments for implementing their projects. Because not every project in the study resulted in all of the above-mentioned benefits, they are considered more ad hoc and should not be expected each time that a motor system is optimized. Of the 41 projects, 22 resulted in reduced maintenance requirements on the motor systems involved. Twelve of these 22 plants were able to quantify their maintenance savings, which together total $900,000. Improvements in productivity in the form of production increases or better product quality were reported in 14 projects, three of which were able to quantify annual revenue increases that total $568,000. Others reported percentage increases in production or decreases in product reject rates. Lower emissions or purchases of ancillary products such as treatment chemicals were reported in eight projects, and two projects resulted in improved plant safety. Also, six projects forestalled equipment purchases and one project averted an expensive asbestos abatement campaign. Together, these seven projects' avoided costs total $770,000. Finally, 10 projects resulted in incentive payments to the firms that

Table 4 Individual project results of 41 total projects

	5 year project life		10 year project life		15 year project life	
	Number of projects	Percentage of projects	Number of projects	Percentage of projects	Number of projects	Percentage of projects
All projects						
Net present value < project cost	27	66	34	83	39	95
IRR > 50%	22	54	24	59	26	63
Simple payback > 2 years	27	66	27	66	27	66
Capital spending						
Net present value > project cost	21	58	28	78	33	92
IRR > 50%	18	50	19	53	20	56
Simple payback < 2 years	21	58	21	58	21	58
Re-engineering						
Net present value > project cost	6	100	6	100	6	100
IRR > 50%	5	83	5	83	6	100
Simple payback < 2 years	6	100	6	100	6	100

performed them that total $1.2 million. When the incentive payments were factored out of the results, the average simple payback for these 10 projects increased by 28% to 6.5, and the IRR of project cash flow decreased by an average of 33% to 46%. When added to the total energy savings achieved, the indirect benefits from motor system optimization projects present a compelling case for their implementation.

Drivers of Motor System Projects

Although each of the projects in the study yielded energy savings, not all of them were implemented with such savings as their primary objective. A review of the reasoning behind each project's implementation revealed a range of drivers for the projects. These motives can be divided into five separate categories: production issues, energy savings, motor system effectiveness, plant expansion, and process reliability. The production issues category refers to projects that were implemented because a motor system was unable to support production processes effectively and the plant was experiencing severe production stoppages and/or high product reject rates. The energy savings group represents plants in which there was recognition that a particular motor system was wasting energy and that optimizing it to capture energy savings was a worthwhile pursuit. The motor system effectiveness label describes motor systems that were able to support production equipment but that operated erratically or unsatisfactorily. The plant expansion category includes plants that underwent an expansion of production equipment and optimized a motor system to support the expansion. Finally, two projects in the study were undertaken with the primary aim of ensuring process reliability. The five categories and the numbers of projects performed according to each are shown in Table 5.

Potential Industry-wide Energy Savings

To provide planning-relevant information useful for strategic planning, the study extrapolates energy savings for U.S. industry based on estimates of annual energy consumption by industrial motor systems. Ballpark estimates of potential industry-wide energy savings were calculated using data from two separate sources: the United States Industrial Motor-Driven Systems Market Assessment, written by the U.S. DOE's EERE, and the EIA's Manufacturing Energy Consumption Survey (MECS) from 1998.

According to EERE's assessment, U.S. industry consumed 691 billion kWh in process motor-driven systems in 1994. The MECS industrial end-use consumption figures reveal that direct industrial uses and processes in U.S. manufacturing plants consumed over 711 billion kWh in 1998. The aggregate pre-project power consumption by the motor systems in the study's 41 plants is 453 million kWh. The energy savings by the plants in the study (106 million kWh) represent a 23% reduction in their aggregate energy consumption. Whether all industrial facilities in the U.S. would achieve similar energy savings rates from implementing projects aimed at improving motor system efficiency is not evident. However, because many of the projects in the study employed best practices in industry for motor system optimization, it is not too far-fetched either. Therefore, a range is presented in Table 6 that indicates how much energy and money could be saved under various implementation rates based on the 23% energy savings rate achieved in this study.

As shown in Table 6, manufacturing plants in the United States could potentially save as much as 83 billion kWh and $4 billion if manufacturing plants that account for 50% of industrial energy consumption would implement projects of similar quality as those in this study. While these sets of figures are tentative, they provide a rough estimate of the range of industry-wide energy savings that industrial motor system optimization projects might achieve in the United States. Additional savings can also be obtained by improving other industrial systems such as steam and direct-fired process heating systems.

Table 5 Drivers of motor system projects

Category	Number of projects	Percent of total
Production issues	8	20
Energy savings	14	34
Motor system effectiveness	15	37
Plant expansion	2	5
Process reliability	2	5

Table 6 Total U.S. energy savings from potential motor system optimization projects (figures are in millions of kilowatt-hours and U.S. dollars)

Energy efficiency and renewable energy assessment			
Implementation rates	10%	25%	50%
Kilowatt-hour	16,228	40,570	81,139
Dollars	779	1,947	3,895
Manufacturing energy consumption survey data			
Implementation rates	10%	25%	50%
Kilowatt-hour	16,698	41,744	83,488
Dollars	801	2,004	4,007

CONCLUSION

This study's intent has been to estimate the impact on industrial competitiveness resulting from motor system optimization projects. The impact of 41 motor system optimization projects was evaluated individually and collectively using three metrics: NPV, IRR, and simple payback. The main findings are that:

- A majority of the 41 projects was successful according to the study's criteria.
- Projects that were implemented in energy intensive manufacturing plants obtained a greater rate of success and a higher rate of return.
- Projects involving large expenditures of capital achieved lower rates of success and return, than projects characterized primarily by engineering or process changes.
- Many projects resulted in positive externalities such as maintenance savings, better productivity, lower emission levels, reduced purchases of ancillary products, improved plant safety, and avoided purchases of plant capital or other costs.

The important finding is that motor system optimization is an underrated source of productivity. Many of the 41 projects were performed in response to production problems. Once the projects were complete, the production problems went away and the plants began to notice the energy savings. Improvements in production occurred in 34% of the plants in this study, and while many of them could not quantify the production impact of their project, they recognized that production increased or that production equipment operated more effectively after the project completion. Instead of re-allocating resources, motor system improvements cause specific systems to use fewer resources (namely, energy) while allowing for a desired production level or standard of quality. Because of this, such projects increase industrial competitiveness and productivity. To view some of the case studies whose data were used in this report, please visit: http://www1.eere.energy.gov/industry/bestpractices/case_studies.html.

BIBLIOGRAPHY

1. Benninga, S. *Financial Modeling*; The MIT Press: Cambridge, MA, 2000; 3–8.
2. Campbell, J.; Lo, A.; Mackinlay, A.C. *The Econometrics of Financial Markets*; Princeton University Press: Princeton, NJ, 1997; 149–152.
3. Clark, W.; Muller, M. *Savings generated by the industrial assessment center program: Fiscal year 2000*. Office of Industrial Productivity and Energy Assessment, 2000; 33.
4. Energy Information Administration. Retail Sales of Electricity, Revenue, and Average Revenue per Kilowatt-hour (and RSEs) by U.S. Electric Utilities to Ultimate Consumers by Census Division, and State, 2000 and 1999—Industrial, 2002.http://www.eia.doe.gov/cneaf/electricity/epav1/ta24p1.html.
5. Scheihing, P.; Rosenberg, M.; Olszewski, M.; Cockrill, C.; Oliver, J. *United states industrial motor-driven systems market assessment: Charting a roadmap to energy savings for industry*. Office of Energy Efficiency and Renewable Energy, 2000; 1–2.
6. Tassey, G. *Lessons learned about the methodology of economic impact studies: The NIST experience*. National Institute of Standards and Technology Strategic Planning and Economic Analysis Group, 1998; 1–4.
7. Xenergy, *United states industrial motor systems market opportunities assessment*. 1998; 2, 10, 49–70.

Insulation: Facilities

Wendell A. Porter
University of Florida, Gainesville, Florida, U.S.A.

Abstract
In this entry, insulation materials and techniques are described in detail. The advantages and disadvantages of the most popular materials and their manufacturing processes are discussed. Information concerning new and emerging techniques and materials is also included. The concept of a building envelope is explained and descriptions of each major building component, such as floors, ceilings and walls, are illustrated in detail.

The proper installation methods for each major type of insulation are explained in both the text and in illustrations. These methods are repeated for the different types of wall construction, frame, masonry, and insulated concrete. The concept of a radiant barrier is also explained, along with related installation instructions.

INTRODUCTION

Insulation is rated in terms of thermal resistance, called *R*-value, which indicates the resistance to heat flow. Although insulation can slow all types of heat flow—conduction, convection, and radiation—its greatest impact is on conduction.

The higher the *R*-value is, the greater the insulation effectiveness is.[1,2] The *R*-value of thermal insulation depends on the type of material, the thickness, and the density. When calculating the *R*-value of a multilayered installation, the *R*-values of the individual layers are added.

The effectiveness of an insulated wall or ceiling also depends on how and where the insulation is installed. For example, compressed insulation will not give its full rated *R*-value. The overall *R*-value of a wall or ceiling will also be somewhat different from the *R*-value of the insulation itself because some heat flows around the insulation through the studs and joists thermal bridging. With careful design, this short-circuiting can be reduced.

The key to an effective insulation system is proper installation of quality insulation products. A building should have a continuous layer of insulation around the entire building envelope (Fig. 1). Studies show that improper installation can cut performance by 30% or more.

INSULATION MATERIALS

The wide variety of insulation materials makes it difficult to determine which products and techniques are the most cost effective (Table 1). Whatever product is chosen, install it per the manufacturer's specifications.

Here are short descriptions of a few of the insulation products available today:

- Fiberglass insulation products come in batt, roll, and loose-fill form, as well as a semirigid board material. Many manufacturers use recycled glass in the production process of fiberglass building insulation, with most using between 20 and 30% recycled glass in their product. Fiberglass is used for insulating virtually every building component—from walls to attics to ductwork.
- The term mineral wool refers to both slag wool and rock wool. Slag wool is manufactured from industrial waste product. It is primarily (~75%) produced from iron ore blast furnace slag, a by-product of smelting. Rock wool is fireproof and produced from natural rocks—basalt primarily—under high heat. Mineral wool insulation is available as a loose-fill product, batts, semirigid, or rigid board. Usage of this product has decreased as more and more building codes require active sprinklering of buildings.
- Cellulose insulation, primarily made from post-consumer recycled newsprint with up to 20% ammonium sulfate and/or borate flame retardants, is installed in loose-fill, wall-spray (damp), dense-pack, and stabilized forms. Because of its high density, cellulose can help reduce air leaks in wall cavities, but air sealing other areas of air infiltration, such as under wall plates and band joists, must be performed to obtain an effective air barrier. However, given certain conditions and applications, cellulose may hold moisture.
- Molded expanded polystyrene (MEPS), often known as beadboard, is a foam product made from molded beads

Keywords: *R*-value; Insulation; Radiant barrier; Ventilation; Batt insulation; Loose-fill insulation; Wall insulation; Floor insulation; Attic insulation.

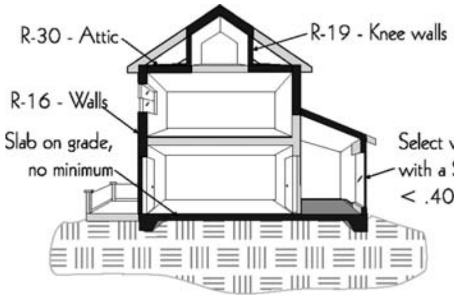


Fig. 1 Building envelope insulation.

of plastic. MEPS is used in several alternative building products discussed in this chapter, including insulated concrete forms and structural insulated panels (SIPs).

- Extruded polystyrene (XPS), also a foam product in rigid board form, is a homogenous polystyrene produced primarily by three manufacturers with characteristic colors of blue, pink, and green.
- Polyisocyanurate, foil-faced rigid board, is insulating foam with one of the highest available *R*-values per inch.
- Closed-cell, high-density spray polyurethane is used both for cavity insulation and as insulating roofing materials [often referred to as spray polyurethane foam (SPF)]. It has structural properties, good adhesive properties, and good compressive strength.

Table 1 Comparison of insulating materials

Material	Typical *R*-value (per inch)
Batts, blankets, and loose-fill insulation	
Mineral wool and fiberglass	2.2–4.0
Cellulose (loose-fill)	3.0–3.7
Cotton (batts)	3.0–3.7
Perlite (loose-fill)	2.5–3.3
Foam insulation and sheathing	
Polyisocyanurate	6.0–6.5
Closed-cell, spray polyurethane	5.8–6.8
Open-cell, low-density polyurethane	3.6–3.8
Extruded polystyrene	5.0
Molded expanded polystyrene (beadboard)	4.0
Fiberboard sheathing (blackboard)	1.3
Air-krete	3.9
OSB sheathing (3/8 in.)	0.5
Foil-faced OSB	Depends on installation
Polyicynene	3.6

Determine actual *R*-values and costs from manufacturers or local suppliers.

- Open-cell, low-density polyurethane foam is used primarily to seal air leaks and provide an insulating layer. Produced primarily from petrochemicals, some of these products are now manufactured in part from soybeans.
- Aerated concrete, including lightweight, autoclaved (processed at high temperature) concrete, can provide a combination of moderate *R*-values and thermal mass for floors, walls, and ceilings, in addition to structural framing.
- Reflective insulation is often used between furring strips on concrete block walls to reflect the heat. Note that reflective insulation products differ from radiant barriers in that they include a trapped air space as part of the product. These trapped air spaces may be a result of the way the reflective insulation is manufactured or installed.

Note that many new types of insulation are rapidly becoming incorporated into conventional construction. However, always research a material's characteristics and suitability to a particular situation before buying any new product. For instance, many new insulation products require covering for fire rating.

Insulation and the Environment

There has been considerable study and debate about the potential negative environmental and health impacts of insulation products.[3] These concerns range from detrimental health effects for the installer to depletion of the earth's ozone layer.

- Fiberglass and mineral wool—questions about effects on health from breathing in fibers. In 2001, the International Agency for Research on Cancer changed its classification for fiberglass and mineral wool from "possible human carcinogen" to "not a known human carcinogen."
- Cellulose—concerns about dust inhalation during installation to VOC emissions from printing inks (these are now almost entirely vegetable-based) and limited evidence of toxicity of boric acid flame retardants. Long-term fire retardancy is unknown. Limited health and safety research has been performed on these products.
- Foam products and chlorofluorocarbons—for years, many foam products contained chlorofluorocarbons (CFCs), which are quite detrimental to the earth's ozone layer. The CFCs were the blowing agent that helped create the lightweight foams. Current blowing agents are:
  - Expanded polystyrene—pentane, which has no impact on ozone layer, but may increase the potential for smog formation.

— Extruded polystyrene, polyisocyanurate, and polyurethane—use primarily hydrochloro-fluorocarbons (HCFCs), which are 90% less harmful to the ozone layer than CFCs. Some companies are moving to non-HCFC blowing agents.
— Open-cell polyurethane, including the products made by Icynene, Inc. and Demilec, Inc., as well as the newer soy-based foams—use water, which is much less detrimental than other blowing agents (Table 2).

INSULATION STRATEGIES

In general, commonly used insulation products are the most economical. Prices can vary according to installer and location. Review all of the choices, as they offer different *R*-values, suggested uses, and environmental and health considerations.

Critical Guidelines

When installing any insulating material, the following guidelines are critical for optimum performance[4]:

- Seal all air leaks between conditioned and unconditioned areas
- Obtain complete coverage of the insulation, especially around doors and windows
- Minimize air leakage through the material with air sealing measures
- Avoid compressing insulation
- Avoid lofting (installing too much air) in loose-fill products

Foam Insulation Strategies

Foam products are primarily economical when they can be applied in thin layers as part of a structural system or to help seal air leaks. Examples include:

- Exterior sheathing over wall framing
- Forms in which concrete can be poured
- As part of a structural insulated panel for building walls
- Spray-applied foam insulation

FLOOR INSULATION

Slab-on-Grade

Slab-on-grade floors consist of a concrete slab poured over at least four inches of compacted gravel or sand and a layer of 10-mil polyethylene used as a vapor barrier. In hot, humid climates, most buildings are built with concrete slab-on-grade.

For colder climates, slabs lose energy as a result of heat conducted outward toward the perimeter of the slab. Insulating the exterior edge of the slab with R-10 rigid insulation can reduce winter heating bills by 10%–20%.

Raised Floor

Raised floor systems (wood and concrete) have specific requirements depending on climate zones. Consult the local building code for specific details.

WALL INSULATION

Walls are the most complex component of the building envelope to provide adequate thermal insulation, air sealing, and moisture control.

Concrete Wall Insulation

Foundation walls and other masonry walls are usually built of concrete block or poured concrete. Insulating concrete block walls is more difficult than insulating framed walls.

Insulating Concrete Block Cores

Builders can insulate the interior cores of concrete block walls with insulation such as:

- Vermiculite R-2.1 per inch (See Fig. 2)
- Polystyrene inserts or beads R-4.0–5.0 per inch
- Polyurethane foam R-5.8–6.8 per inch

Unfortunately, as shown in Fig. 2, the substantial thermal bridging in the concrete connections between cores continues to depreciate the overall *R*-value. This approach is only a partial solution to providing a quality, well-insulating wall. Other techniques, as explained in the next few pages, provide more cost-effective solutions.

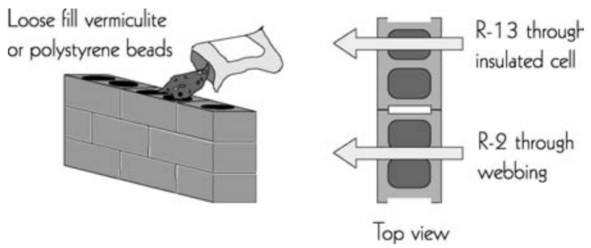


Fig. 2 Insulating concrete block cores (R-4–R-6 overall).

Table 2 Comparison of insulation materials (environmental characteristics and other information)

Type of insulation	Installation method(s)	*R*-value per inch (RSI/m)[a]	Raw materials	Pollution from manufacture	Indoor air quality impacts	Comments
Fiber insulation						
Cellulose	Loose fill; wall-spray (damp); dense-pack; stabilized	3.0–3.7 (21–26)	Old newspaper, borates, ammonium sulfate	Vehicle energy use and pollution from newspaper recycling	Fibers and chemicals can be irritants. Should be isolated from interior space	High recycled content; very low embodied energy
Fiberglass	Batts; loose fill; semi-rigid board	2.2–4.0 (15–28)	Silica sand; limestone; boron; recycled glass, phenol formaldehyde resin or acrylic resin	Formaldehyde emissions and energy use during manufacture; some manufactured without formaldehyde	Fibers can be irritants, and should be isolated from interior spaces. Formaldehyde is a carcinogen. Less concern about cancer from respirable fibers	
Mineral wool	Loose fill; batts; semi-rigid or rigid board	2.8–3.7 (19–26)	Iron-ore blast furnace slag; natural rock; phenol formaldehyde binder	Formaldehyde emissions and energy use during manufacture	Fibers can be irritants, and should be isolated from interior spaces. Formaldehyde is a carcinogen. Less concern about cancer from respirable fibers	Rigid board (e.g., Roxul) can be an excellent foundation drainage and insulation material
Cotton	Batts	3.0–3.7 (21–26)	Cotton and polyester mill scraps (especially denim)	Negligible	Considered very safe	Two producers; also used for flexible duct insulation
Perlite	Loose fill	2.5–3.3 (17–23)	Volcanic rock	Negligible	Some nuisance dust	

[a]RSI/m: The standard unit of measurement in the United States has been the Imperial unit. The country is converting to the International System (SI) unit—or metric standard—which predominates internationally. To differentiate like terms, you may find "SI" added to the term symbol. For example, RSI refers to the *R*-value in International System (metric) units.

Conversion factors

R-value conversions	To get	Multiply	By
Thermal resistance (*R*)	RSI (m^2 C/w)	R (ft^2 hF/Btu)	0.1761
Insulation *R*/unit thickness	RSI/mm	*R*/in.	0.00693
In the chart, the heading is *R*-value per inch (RSI/m); to obtain this number, the RSI/mm is divided by 1000.			

(Continued)

Table 2 Comparison of insulation materials (environmental characteristics and other information) (*Continued*)

Type of insulation	Installation method(s)	*R*-value per inch (RSI/m)[a]	Raw materials	Pollution from manufacture	Indoor air quality impacts	Comments
Foam insulation						
Polyisocyanurate	Foil-faced rigid boards; nail-base with OSB sheathing	6.0–6.5 (42–45)	Fossil fuels; some recycled PET; pentane blowing agent; TCPP flame retardant; aluminum facing	Energy use during manufacture	Potential health concerns during manufacture. Negligible emissions after installation	Phaseout of HCFC ozone-depleting blowing agents completed
Extruded polystyrene (XPS)	Rigid board	5.0 (35)	Fossil fuels; HCFC-142b blowing agent; HBCD flame retardant	Energy use during manufacture. Ozone depletion	Potential release of residual styrene monomer (a carcinogen) and HBCD flame retardant	Last remaining insulation material with ozone-depleting blowing agents
Expanded polystyrene (EPS)	Rigid board	3.6–4.4 (25–31)	Fossil fuels; pentane blowing agent; HBCD flame retardant	Energy use during manufacture	Potential release of residual styrene monomer (a carcinogen) and HBCD flame retardant	
Closed-cell spray polyurethane	Spray-in cavity-fill or spray-on roofing	5.8–6.8 (40–47)	Fossil fuels, HCFC-141b (through early 2005) or HFC-245fa blowing agent; nonbrominated flame retardant	Energy use during manufacture, global-warming potential from HFC blowing agent	Quite toxic during installation (respirators or supplied air required). Allow several days of airing out prior to occupancy	
Open-cell, low-density polyurethane	Spray-in cavity-fill	3.6–3.8 (25–27)	Fossil fuels and soybeans; water as a blowing agent; nonbrominated flame retardant	Energy use during manufacture	Quite toxic during installation (respirators or supplied air required). Allow several days of airing out prior to occupancy	
Air-Krete	Spray-in cavity-fill	3.9 (27)	Magnesium oxide from seawater; ceramic talc	Negligible	Considered very safe	Highly fire-resistant; inert; remains friable
Radiant barrier						

Bubble back	Stapled to framing	Depends on installation	Aluminum; fossil fuels	Energy use during manufacture	Minimal offgassing from plastic	Exaggerated *R*-value claims have been common
Foil-faced polyethylene foam	Stapled to framing; requires air space for radiant benefit	Depends on installation	Aluminum; fossil fuels; recycled polyethylene	Energy use during manufacture	Minimal offgassing from polyethylene	Exaggerated *R*-value claims have been common. Recycled content in some
Foil-faced paperboard sheathing	Stapled to framing; requires air space for radiant benefit	Depends on installation	Aluminum; fossil fuels; recycled paper	Energy use during manufacture	Considered very safe	High recycled content. Structural sheathing available (e.g., Thermo-Ply®)
Foil-faced OSB	Most common as attic sheathing	Depends on installation	Wood fiber; formaldehyde binder in OSB; aluminum	Energy use and VOC emissions during manufacture	Formaldehyde emissions	Primary benefit is reduced heat gain

Exterior Rigid Fiber Glass or Foam Insulation

Rigid insulation is generally more expensive per *R*-value than mineral wool or cellulose, but its rigidity is a major advantage (Fig. 3). However, it is difficult and expensive to obtain *R*-values as high as in framed walls.

Interior Foam Wall Insulation

Foam insulation can be installed on the interior of concrete block walls (Fig. 4); however, it must be covered with a material that resists damage and meets local fire code requirements. Half-inch drywall will typically comply, but furring strips will need to be installed as nailing surfaces. Furring strips are usually installed between sheets of foam insulation; however, to avoid the direct, uninsulated thermal bridge between the concrete wall and the furring strips, a continuous layer of foam should be installed underneath or on top of the nailing strips.

Interior Framed Wall

In some cases, designers will specify a framed wall on the interior of a masonry wall (Fig. 5). Standard framed wall insulation and air-sealing practice can then be applied.

Insulated Concrete Form Systems

Insulated concrete forms (ICFs) are permanent rigid plastic foam forms that are filled with reinforced concrete to create structural walls with significant thermal insulation (Fig. 6). The foam is typically either expanded polystyrene (EPS) or extruded polystyrene (XPS) and occasionally polyurethane, but it may also be made from a composite of cement and foam insulation or a composite of cement and processed wood.[5]

The concrete will be one of several shapes: flat, waffle- or screen-grid, or post-and-beam, depending on the specific form design. The Portland Cement Association (PCA) reports that in 1994, 0.1% of all new homes used ICFs in above-grade walls (about 1100 new homes). That number rose to 1.2% in 1999, 2.7% in 2001, and increased to 3.8% in 2002, which would be 50,639 homes, according to U.S. Census Data.

Above-grade ICF walls cost more to build than typical wood-framed walls. As wood-framed walls approach the thermal insulation value of ICFs, the cost differential will decrease. In most cases, materials' costs (concrete and forms) are primarily responsible for increased costs, while labor costs are often similar to wood framing. Cost premium depends on relative material prices, labor efficiency for each system, engineering necessity, and its effect on other practices or trades, among other factors.

The cost premium for ICF houses is smaller in areas such as high-wind regions, which require additional labor, time, and materials for special construction of

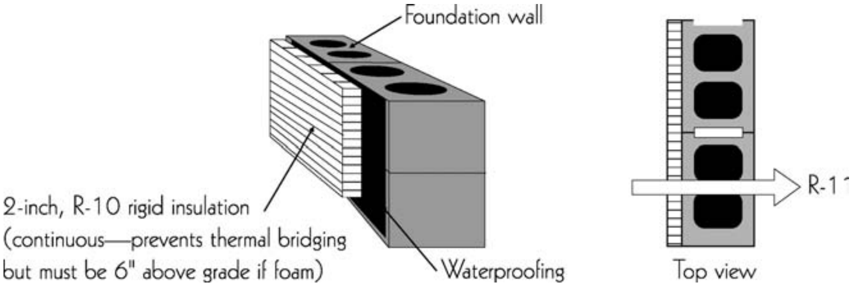


Fig. 3 Exterior foam insulation (R-11–R-12 overall).

wood-framed houses. According to an NAHB Research Center study, costs are estimated to increase by 1%–8% of total house cost over a wood-framed house.

Lightweight Concrete Products

Lightweight, air-entrained concrete is an alternative wall system (Fig. 7). Autoclaved aerated concrete (AAC), sometimes referred to as precast autoclaved aerated concrete (PAAC), which can be shipped either as blocks or panels, combines elevated *R*-values (compared to standard concrete) with thermal mass.

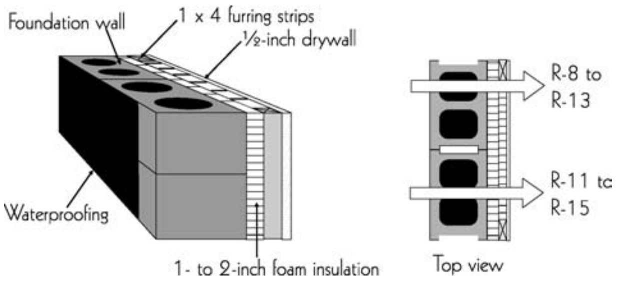


Fig. 4 Interior foam insulation (R-10–R-14 overall).

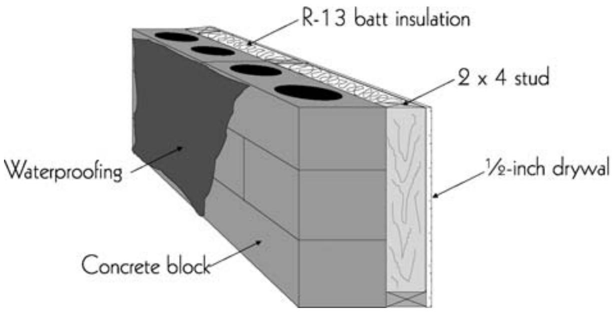


Fig. 5 Interior framed wall insulation (R-11–R-13 overall).

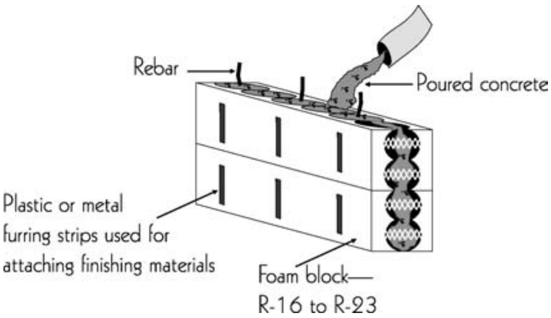


Fig. 6 Insulated concrete foam system (R-17–R-24 overall).

2×4 Wall Insulation

Throughout the United States, debates on optimal wall construction continue.[6] Table 3 summarizes typical problems and solutions in walls framed with 2×4 studs. In addition to standard framing lumber and fasteners, the following materials will also be required during construction:

- Foam sheathing for insulating headers.
- 1×4 or metal T-bracing for corner bracing (Fig. 8).
- R-13 batts for insulating areas during framing behind shower/tub enclosures and other hidden areas.
- From the Florida Building Code, Residential (FBC-R): "R307.2 Bathtub and shower spaces. Bathtub and shower floors and walls above bathtubs with installed shower heads and in shower compartments shall be finished with a nonabsorbent surface. Such wall surfaces shall extend to a height of not less than 6 ft above the floor."
- Caulking or foam sealant for sealing areas that may be more difficult to seal later.

Avoid Side Stapling

Walls are usually insulated with batts that have attached vapor retarder facing. Many builders question whether it is best to side staple or face staple batt insulation. The common arguments are that face stapling results in less

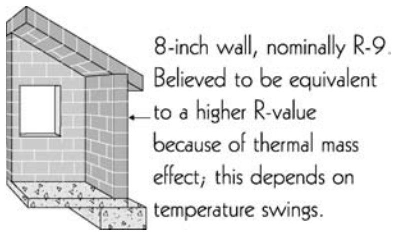


Fig. 7 Lightweight concrete products (R-1.1 per inch plus).

Table 3 2×4 Framed wall problems and solutions

Problem	Solution
Small space available for insulation	Install continuous exterior foam sheathing and medium (R-13) to high (R-15) density cavity insulation
Enclosed cavities are more prone to cause condensation, particularly when sheathing materials with low *R*-values are used	Install a continuous air barrier system. Use continuous foam sheathing
Presence of wiring, plumbing, ductwork, and framing members lessens potential *R*-value and provides pathways for air leakage	Locate mechanical systems in interior walls; avoid horizontal wiring runs through exterior walls; use air sealing insulation system

compression, while side stapling interferes less with drywall installation.

The ideal solution should focus on where the kraft paper (vapor retarder) is rather than on how it is installed.[7]

The face stapling question is an appropriate question in northern or "heating-dominated" climates. In northern areas, vapor retarders should be installed on the "warm" side of the wall cavity. In southern or "cooling-dominated" climates, the vapor retarder should be on the outside surface of the wall cavity. Because of this, the use of unfaced batts is recommended in hot-humid climates (Fig. 9).

Unfaced batts are slightly larger than the standard 16- or 24-inch stud spacing and rely on a friction-fit for support. Because unfaced batts are not stapled, they can often be installed in less time. In addition, it is easier to cut unfaced batts to fit around wiring, plumbing, and other obstructions in the walls.

Blown Loose-Fill Insulation

Loose-fill cellulose, fiberglass, and rock wool insulation can also be used to insulate walls. This insulation is installed with a blowing machine and held in place with a glue binder or netting (Fig. 10). This technique can provide good insulation coverage in the stud cavities; however, it is very important that excess moisture in the binder be allowed to evaporate before the wall cavities are enclosed by an interior finish. Keep in mind that insulation products are not replacements for proper air sealing technique.

Blown Foam Insulation

Some insulation contractors are now blowing polyurethane or polyicynene insulation into the walls and ceilings of new buildings (Fig. 11). This technique provides high *R*-values in relatively thin spaces and seals air leaks effectively. The economics of foam insulation should be examined carefully prior to their application.

Structural Insulated Panels

Another approach to wall construction is the use of structural insulated panels (SIP), also known as stress-skin panels (Fig. 12).[8] They consist of 4-inch or 6-inch thick

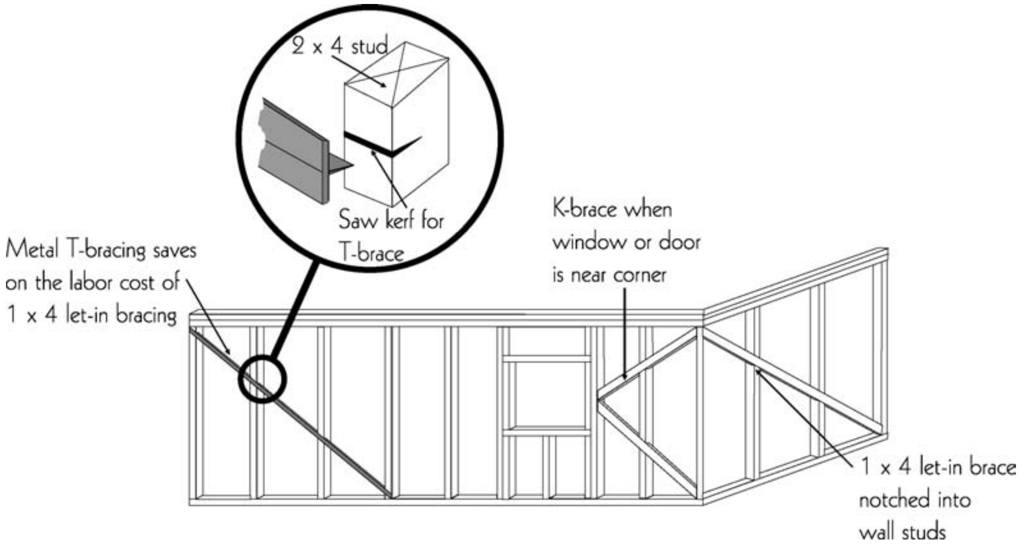


Fig. 8 Let-in bracing.

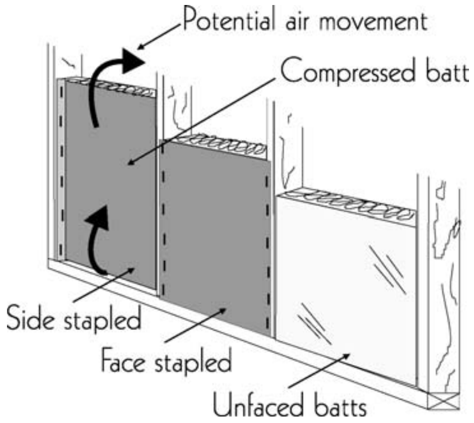


Fig. 9 Insulating walls with batts.

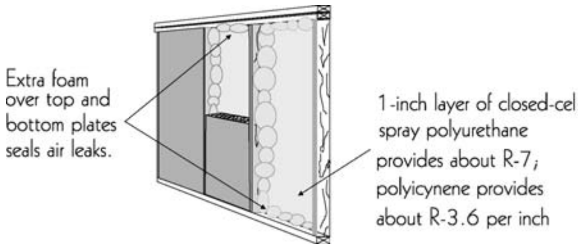


Fig. 11 Blown foam insulation.

foam panels onto which structural sheathing, such as oriented strand board (OSB), cement fiber board, or various types of metal have been attached. They reduce labor costs, and because of the reduced framing in the wall, they have higher *R*-values and less air leakage than standard walls.

SIPs are generally four feet wide and eight to twelve feet long. There are a wide variety of manufacturers, each with its own method of attaching panels together. Procedures for installing windows, doors, wiring, and plumbing have been worked out by each manufacturer. Some SIPs come from the factory with preinstalled windows. In addition to their use as wall framing, SIPs can also be used in ceilings, floors, and roofs.

While buildings constructed with SIPs may be more expensive than those with standard framed and insulated walls, research studies have shown that SIP-built buildings have higher average insulating values per inch than most commonly used insulation materials. Due to their typical modular style of construction, infiltration losses are also reduced. Thus, they can provide substantial energy savings. Be sure to follow local building codes with regard to termites, including leaving a 6” inspection zone above finished grade.

The performance of any SIP wall depends on its component materials and installation processes. There are a few important variables to take into consideration when building with SIP systems:

1. Panel fabrication (proper panel gluing, pressing, and curing) is critical to prevent delamination.
2. Panels must be flat, plumb, and have well-designed connections to ensure tightness of construction.
3. Though SIPs offer ease of construction, installers may need training in installing the system being used.
4. Fire rating of SIP materials and air-tightness of SIP installation affect the system’s fire safety.
5. There may be potential insect and rodent mitigation issues, depending on SIP materials and construction.
6. Proper HVAC design and installation must take the SIP system being used into account.

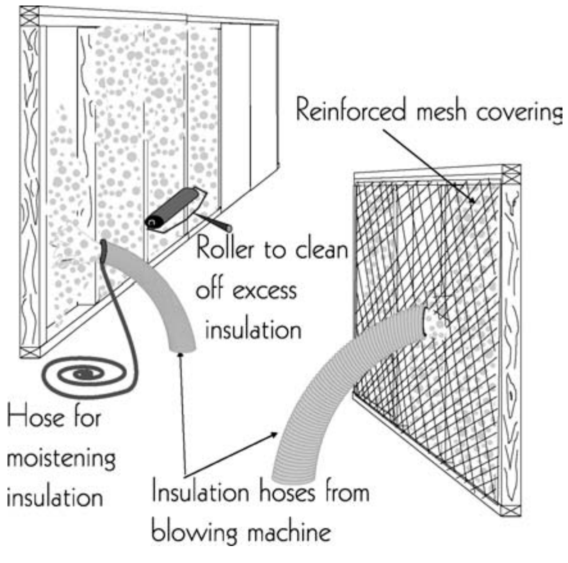


Fig. 10 Blown sidewall insulation options.

Metal Framing

Builders and designers are well aware of the increasing cost and decreasing quality of framing lumber. As a consequence, interest in alternative framing materials,

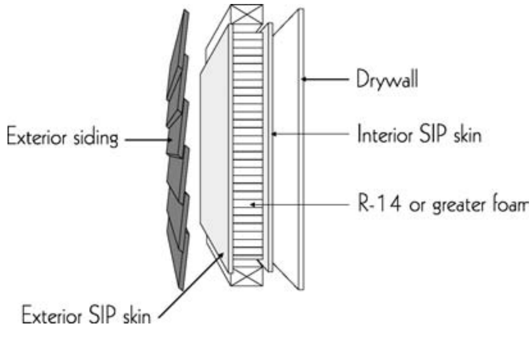


Fig. 12 Structural insulated panels (SIP).

such as metal framing, has grown. While metal framing offers advantages over wood, such as consistency of dimensions, lack of warping, and resistance to moisture and insect problems, it has distinct disadvantages from an energy perspective.

Metal framing is an excellent conductor of heat. Buildings framed with metal studs and plates usually have metal ceiling joists and rafters, as well. Thus, the entire structure serves as a highly conductive thermal grid. Insulation placed between metal studs and joists is much less effective due to the extreme thermal bridging that occurs across the framing members.

The American Iron and Steel Institute is well aware of the challenges involved in building an energy efficient steel structure. In their publication *Thermal Design Guide for Exterior Walls (Publication RG-9405)*, the Institute provides information on the thermal performance of steel-framed buildings. Table 4 summarizes some of their findings.

Moisture-related problems have been reported in metal frame buildings that do not use sufficient insulated sheathing on exterior walls. Metal studs cooled by the air conditioning system can cause outdoor air to condense, leading to mildew streaks (or ghosting), where one can see the framing members on the inside and outside of a home. In winter, studs covered by cold outside air can also cause streaking. Attention to proper insulation techniques can alleviate this problem.

2×6 Wall Construction

There has been interest in hot, humid climates in the use of 2×6s for construction. The advantages of using wider wall framing are:

- More space provides room for R-19 or R-21 wall insulation.
- Thermal bridging across the studs is less of a penalty due to the higher *R*-value of 2×6s.
- Less framing reduces labor and material costs.
- There is more space for insulating around piping, wiring, and ductwork.

Table 4 Effective steel wall *R*-values

Cavity insulation	Sheathing	Effective overall *R*-value
11	2.5	9.5
11	5	13
11	10	18
13	2.5	10
13	5	14

Disadvantages of 2×6 framing include:

- Wider spacing may cause the interior finish or exterior siding to bow slightly between studs.
- Window and door jambs must be deeper, resulting in additional costs.
- Walls with substantial window and door area may require almost as much framing as 2×4 walls and leave relatively little area for actual insulation.

The economics of 2×6 wall insulation are affected by the number of windows in the wall because each window opening adds extra studs and may require the purchase of a jamb extender. Walls built with 2×6s having few windows provide positive economic payback. However, for walls in which windows make up over 10% of the total area, the economics become questionable.

CEILINGS AND ROOFS

Attics over flat ceilings are usually the easiest part of a building's exterior envelope to insulate. They are accessible and have ample room for insulation. However, many homes use cathedral ceilings that provide little space for insulation.[9] It is important to insulate both types of ceilings properly.

Attic Ventilation

In the summer, properly designed ventilation reduces roof and ceiling temperatures, thus potentially saving on cooling costs and lengthening the life of the roof. In winter, roof vents expel moisture which could otherwise accumulate and deteriorate insulation or other building materials.

At present, several research studies are investigating whether attic ventilation is beneficial. For years, researchers have believed the cooling benefits of ventilating a well-insulated attic to be negligible. However, some experts are now questioning whether ventilation is even effective at moisture removal. The Florida Building Code, Residential now provides provisions for "conditioned attic assemblies" (Section R806.4) as long as certain conditions are met. When attic ventilation is provided, ventilation openings shall be provided with corrosion-resistant wire mesh with 1/8 in. minimum to 1/4 in. maximum openings. Total net free ventilation area shall not be less than 1–150 of the area of the space ventilated. An exception for 1–300 is provided in the code.

Vent Selection

The amount of attic ventilation needed is based on state building code requirements. If ventilating the roof, locate vents high along the roof ridge and low along the eave or

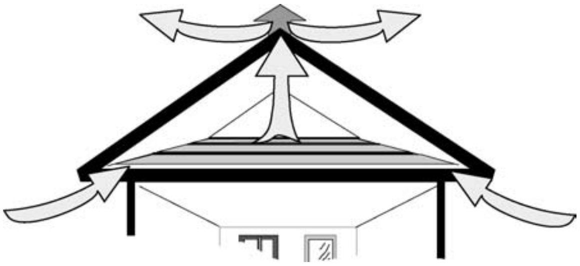


Fig. 13 Attic ventilation through soffit and ridge Vents.

soffit. Vents should provide air movement across the entire roof area (Fig. 13). There are a wide variety of products available, including ridge, gable, soffit, and mushroom vents.

To allow for proper airflow in attic spaces, it is common practice to install a rafter baffle at the soffit. This will prevent insulation from sealing off the airflow from the soffit vent to the attic space.

The combination of continuous ridge vents along the peak of the roof and continuous soffit vents at the eave provides the most effective ventilation. Ridge vents come in a variety of colors to match any roof. Some brands are made of plastic and covered by cap shingles to hide the vent from view.

Manufacturer or product testing is being performed by a variety of organizations to verify leak-free operation of continuous ridge vents in high wind situations. Care should be taken to ensure that the vents chosen are appropriate for hurricane-prone areas.

Powered Attic Ventilator

Electrically powered roof ventilators can consume more electricity to operate than they save on air conditioning costs and are not recommended for most designs. Power vents can create negative pressures in the home, which may have detrimental effects, such as (Fig. 14):

- Drawing outside air into the home
- Removing conditioned air from the home through ceiling leaks and bypasses
- Pulling pollutants, such as radon and sewer gases, into the home
- Backdrafting fireplaces and fuel-burning appliances

Attic Floor Insulation Techniques

Either loose-fill or batt insulation can be installed on an attic floor. Generally, blowing loose-fill attic insulation is usually less expensive than installing batts or rolls. Most attics have either blown fiberglass, rock wool, or cellulose. Ceilings with a rise greater than 5 and a run of 12 (5 over 12) should not be insulated with blown-in insulation.

Loose-Fill Attic Insulation

Steps for installing loose-fill attic insulation[10]:

1. Seal attic air leaks, as prescribed by fire and energy codes.
2. Follow manufacturer's and state building code clearance requirements for heat-producing equipment found in an attic, such as flues or exhaust fans. One example of attic blocking is shown later in this chapter.
3. Use baffles to preserve ventilation space at eave of roof for soffit vents.
4. Insulate the attic hatch or attic stair. Foam boxes are available for providing a degree of insulation over a pull-down attic stairway.
5. Determine the attic insulation area; based on the spacing and size of the joists, use the chart on the

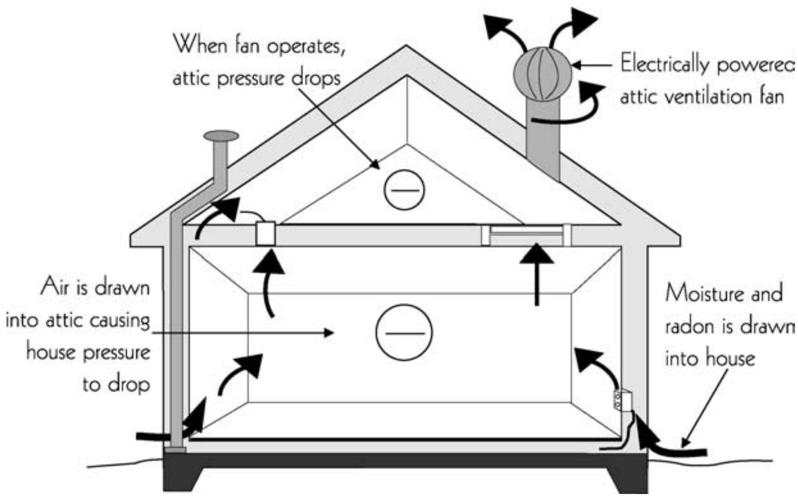


Fig. 14 Attic ventilation through powered ventilation.

Table 5 Blown cellulose in attics

			2×6 joists spaced 24 in. on center		2×6 joists spaced 16 in. on center	
***R*-value at 75°F**	**Minimum thickness (in.)**	**Minimum weight (lb/ft^2)**	**Coverage per 25-lb bag (ft^2)**	**Bags per 1000 ft^2**	**Coverage per 25-lb bag (ft^2)**	**Bags per 1000 ft^2**
R-40	10.8	2.10	12	83	13	77
R-32	8.6	1.60	16	63	18	56
R-24	6.5	0.98	21	48	23	43
R-19	5.1	0.67	37	27	41	24

insulation bag to determine the number of bags to install. Table 5 shows a sample chart for cellulose insulation. Cellulose is heavier than fiberglass for the same *R*-value. Closer spacing of roof joists and thicker drywall is required for larger *R*-values. Check this detail with the insulation contractor. Weight limits and other factors at R-38 insulation levels are shown in Table 5 for the three primary types of loose fills.

6. Avoid fluffing the insulation (blowing with too much air) by using the proper air-to-insulation mixture in the blowing machine. A few insulation contractors have "fluffed" (added extra air to) loose-fill insulation to give the impression of a high *R*-value. The insulation may be the proper depth, but if too few bags are installed, the *R*-values will be less than claimed.
7. Obtain complete coverage of the blown insulation at relatively even insulation depths. Use attic rulers (obtainable from insulation contractors) to ensure uniform depth of insulation.

Batt Attic Insulation

Steps for installing batt insulation:

1. Seal attic air leaks, as prescribed by fire and energy codes.
2. Block around heat-producing devices, as described in Step 2 for loose-fill insulation.
3. Insulate the attic hatch or attic stair as described in Step 4 for loose-fill insulation.
4. Determine the attic insulation area based on the spacing and size of the joists, order sufficient R-30 insulation for the flat attic floor. Choose batts that are tapered—cut wider on top—so that they cover the top of the ceiling joists. (See Fig. 15).
5. When installing the batts, make certain they completely fill the joist cavities. Shake batts to ensure proper loft. If the joist spacing is uneven, patch gaps in the insulation with scrap pieces. Try not to compress the insulation with wiring, plumbing, or ductwork. In general, obtain complete coverage of full-thickness, noncompressed insulation.
6. Attic storage areas can pose a problem. If the ceiling joists are shallower than the depth of the insulation (generally less than 2×10s), raise the finished floor using 2×4s or other spacing lumber. Install the batts before nailing the storage floor in place (see Fig. 16).

Note that often attic framing is not designed for storage. Check engineered loads of framing before increasing loads and piggy-backing ceiling joists.

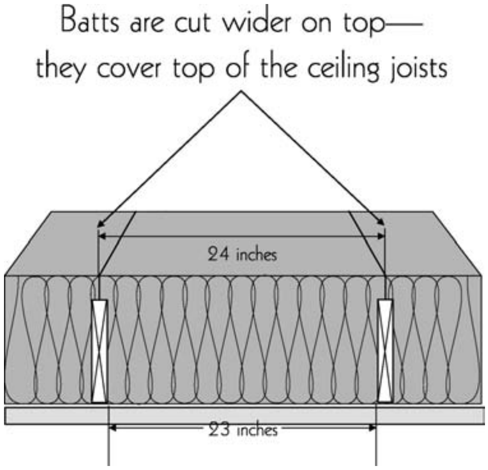


Fig. 15 Full-with ceiling batt insulation.

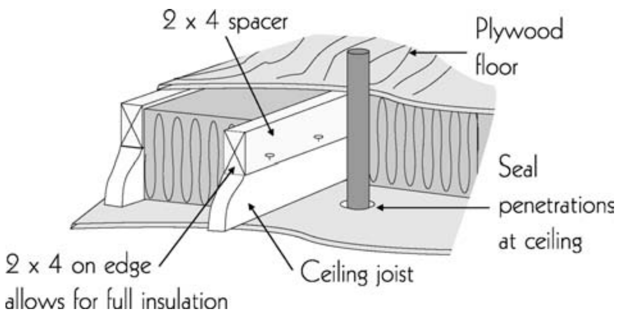


Fig. 16 Ceiling insulation under attic storage floor.

Indu–Inter

Preventing Air Flow Restrictions at the Eave

One problem area in many standard roof designs is at the eave, where there is not enough room for full R-30 insulation without preventing air flow from the soffit vent or compressing the insulation, which reduces its *R*-value. Figs. 17 and 18 show several solutions to this problem. If using a truss roof, purchase raised heel trusses that form horizontal overhangs. They should provide adequate clearance for both ventilation and insulation.

In stick-built roofs, where rafters and ceiling joists are cut and installed on the construction site, an additional top plate that lies across the top of the ceiling joists at the eave will prevent compression of the attic insulation. Note: This needs to be a double plate for bearing unless rafters sit directly above joists. The rafters sitting on this raised top plate allow for both insulation and ventilation.

Cathedral Ceiling Insulation Techniques

Cathedral ceilings are a special case because of the limited space for insulation and ventilation within the depth of the rafter. Fitting in a 10-in. batt (R-30) and still providing ventilation is impossible with a 2×6 or 2×8 rafter (R-19 or R-25, respectively).

Building R-30 Cathedral Ceilings

Cathedral ceilings built with 2×12 rafters can be insulated with standard R-30 batts and still have plenty of space for ventilation. Some builders use a vent baffle between the insulation and roof decking to ensure that the ventilation channel is maintained.

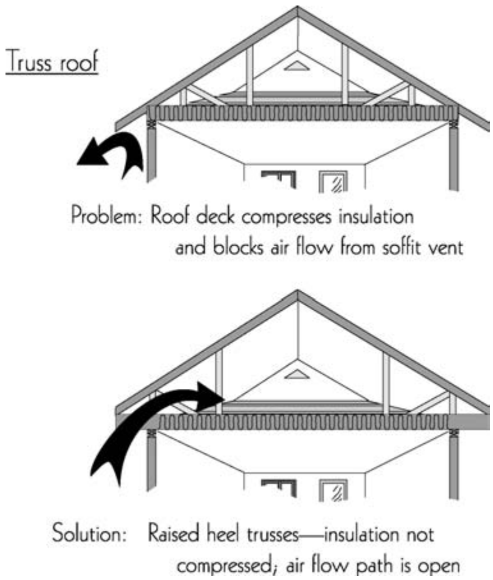


Fig. 17 Soffit air ventilation—raised heel trusses.

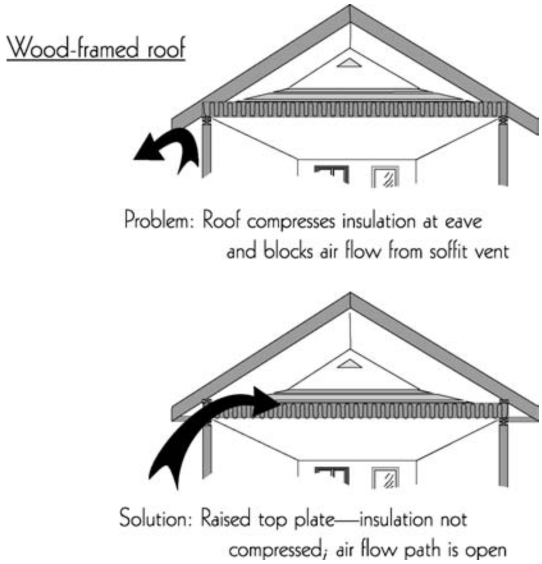


Fig. 18 Soffit air ventilation—raised top plate.

If 2×12s are not required structurally, most builders find it cheaper to construct cathedral ceilings with 2×10 rafters and high-density R-30 batts, which are 8¼ in. thick (Table 6).

Some contractors wish to avoid the higher cost of 2×10 lumber and use 2×8 rafters. These roofs are usually insulated with R-19 batts.

In framing with 2×6 and 2×8 rafters, higher insulating values can be obtained by installing rigid foam insulation under the rafters. Note that the rigid foam insulation must be covered with a fire-rated material when used on the interior of the building. Drywall usually meets this requirement.

Scissor Trusses

Scissor trusses are another cathedral ceiling framing option. Make certain they provide adequate room for both R-30 insulation and ventilation, especially at their ends, which form the eave section of the roof.

Difficulties with Exposed Rafters

A cathedral ceiling with exposed rafters or roof decking is difficult and expensive to insulate well. Often, foam

Table 6 Cathedral ceiling insulation options

Rafter	Batt
2×8	R-19
2×10	R-25
2×10	Moderate density R-30
2×12	Standard density R-30

Any sized rafter; blown-in cellulose, fiberglass, or rock wool held in place; provide 1 in. ventilation space above.

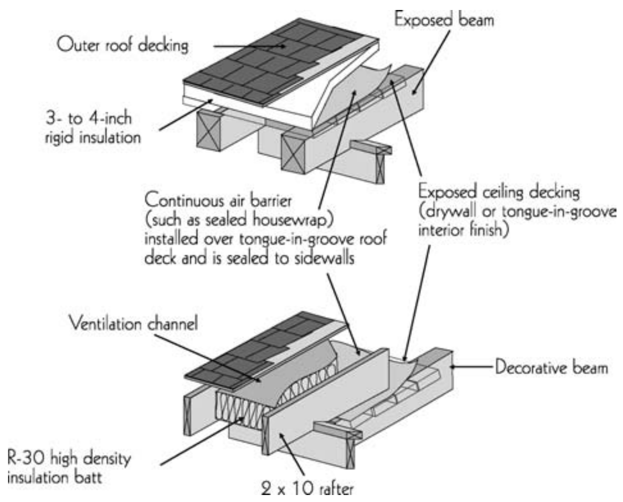


Fig. 19 Insulating exposed rafters.

insulation panels are used over the attic deck, as shown in Fig. 19. However, to achieve R-30, four to seven inches of foam insulation are needed. Ventilation is also a problem.

In homes where exposed rafters are desired, it may be more economical to build a standard, energy efficient cathedral ceiling, and then add exposed decorative beams underneath. Note that homes having tongue-and-groove ceilings can experience substantially more air leakage than solid, drywall ceilings. Install a continuous air barrier, sealed to the walls above the tongue-and-groove roof deck and held in place; provide 1 in. ventilation space above.

Recessed Lights

Standard recessed fixtures require a clearance of several inches between the sides of the lamp's housing and the attic insulation. In addition, insulation cannot be placed over the fixture. Even worse, recessed lights leak considerable air between attics and the home.

Insulated ceiling (IC) rated fixtures have a heat sensor switch that allows the fixture to be covered—except for the top—with insulation (see Fig. 20 for the proper insulation methods for these fixtures). However, these units also leak air. If you have to use recessed lights, install airtight IC-rated fixtures. There are alternatives to recessed lights, including surface-mounted ceiling fixtures and track lighting, which typically contribute less air leakage to the home.

RADIANT HEAT BARRIERS

Radiant heat barriers (RHB) are reflective materials that can reduce summer heat gain via the insulation and building materials in attics and walls. RHBs work two ways: first, they reflect thermal radiation well; and, second, they emit (give off) very little heat. RHBs should always face a vented airspace and be installed to prevent dust build-up. They are usually attached to the underside of the rafter or truss top chord or to the underside of the roof decking. Acceptable attic radiant barrier configurations can be found in Fig. 21.

How Radiant Barrier Systems Work

A radiant barrier reduces heat transfer. Thermal radiation, or radiant heat, travels in a straight line away from a hot surface and heats any object in its path.

When sunshine heats a roof, most of the heat conducts through the exterior roofing materials to the inside surface

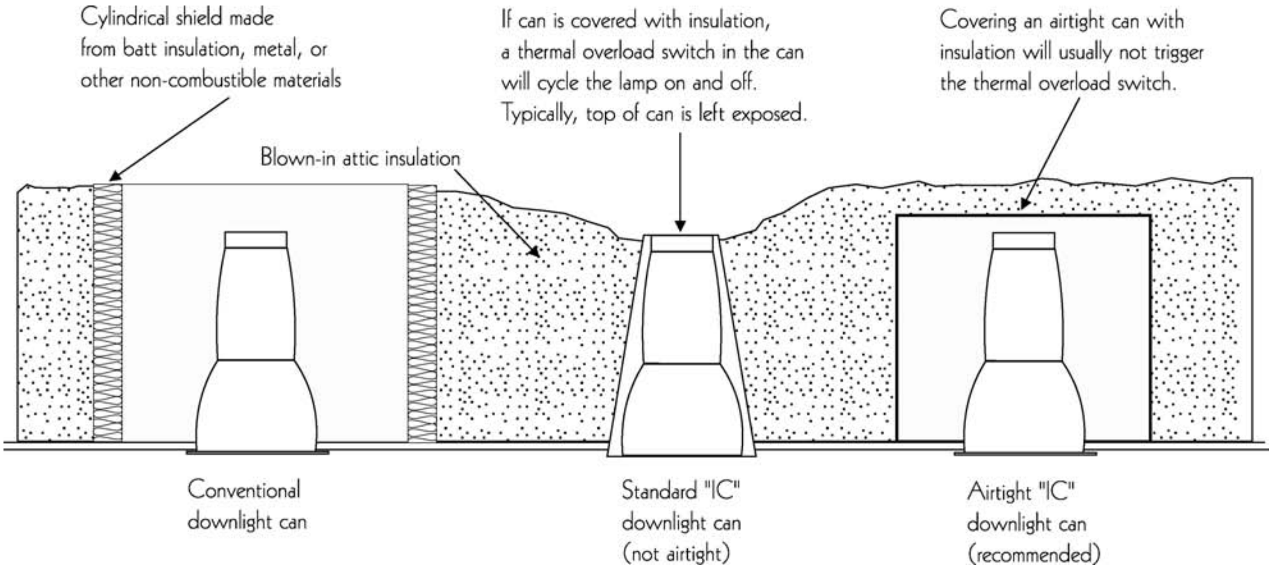


Fig. 20 Recessed lighting insulation.

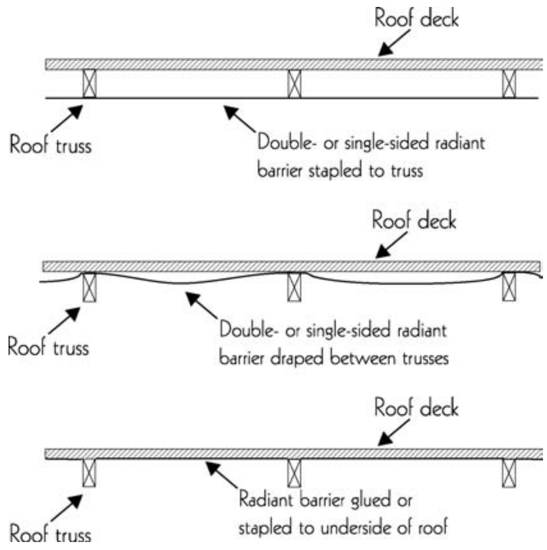


Fig. 21 Radiant Barrier Configuration.

of the roof sheathing. Heat then transfers by radiation across the attic space to the next material—either the top of the attic insulation or the attic floor. A radiant barrier, properly installed in one of many locations between the roof surface and the attic floor, will reduce radiant heat flow. Thermal insulation on the attic floor resists the flow of heat through the ceiling into the living space below. The rate at which insulation resists this flow determines the insulation's *R*-value. The amount of thermal insulation affects the potential radiant barrier energy savings. For example, installing a radiant barrier in an attic that already has high levels of insulation (R-30 or above) would result in much lower energy savings than an attic insulated at a low level (R-11 or less).

All radiant barriers use reflective foil that blocks radiant heat transfer. In an attic, a radiant barrier that faces an air space can block up to 95% of the heat radiating down from a hot roof. Only a single, thin, reflective surface is necessary to produce this reduction in radiant heat transfer. Additional layers of foil do little more to reduce the remaining radiant heat flow.

Conventional types of insulation consist of fibers or cells that trap air or contain a gas to retard heat conduction. These types of insulation reduce conductive and radiant heat transfer at a rate determined by their *R*-value, while radiant barriers reduce only radiant heat transfer. There is no current method for assigning an *R*-value to radiant barriers. The reduction in heat flow achieved by the installation of a radiant barrier depends on a number of factors, such as ventilation rates, roof reflectivity, ambient air temperatures, geographical location, the amount of roof solar gains, and the amount of conventional insulation present.

Several factors affect the cost effectiveness of installing a radiant barrier. You should examine the performance and cost savings of at least three potential insulation options: adding additional conventional insulation, installing a radiant barrier, and adding both conventional insulation and a radiant barrier.

In 1991 (revised June, 2001), the U.S. Department of Energy (DOE) published the Radiant Barrier Attic Fact Sheet, which shows how to calculate the economics of radiant barriers and added ceiling insulation. It includes an Energy Savings Worksheet with an example. The worksheet is part of the fact sheet and can be found at http://www.ornl.gov/sci/roofs+walls/radiant/rb_05.html.[11]

Because radiant barriers redirect radiant heat back through the roofing materials, shingle temperatures may increase between 1 and 10°F (0.6°C–5.6°C). This increase does not appear to exceed the roof shingle design criteria. The overall effect on roof life, if any, is not known.

Remember, radiant barriers are most effective in blocking summer radiant heat gain and saving air-conditioning costs. Although the radiant barrier may be somewhat effective in retaining heat within a cold-climate home, it may also block any radiant winter solar heat gain in the attic.

CONCLUSION

Buildings are insulated to help moderate the environment that we live and work in. As utility costs rise, this aspect of our constructed environment becomes more and more important. Many new materials have entered the market in the past few years, each having their own particular advantages and disadvantages. This article has described in detail the importance of installation practices and their effect on overall performance. Emphasis was placed on properly defining the building envelope and different methods and materials that can be used to effectively insulate and, therefore, thermally isolate the constructed environment from the daily extremes produced by local weather.

New techniques and materials will evolve; however, some guidelines will remain the same:

- Choose an insulation material with characteristics suitable for the climate region.
- Use the proper *R*-value for the climate region where each particular structure is being constructed.
- Apply the recommended thickness, or *R*-value, to each building component, such as ceiling, walls, and floors.
- Install each particular insulation product in a proper fashion, following the manufacturer's recommendations and instructions, leaving no voids, and filling each building cavity to the level required for the building site.

Insulation materials and techniques play a large part in minimizing energy consumption and maximizing human comfort. Recent advances in both materials and techniques will help further this trend.

ACKNOWLEDGMENTS

Many thanks to Kyle Allen, an engineering student at the University of Florida, because without his persistent efforts, the publication of this article would not have been possible.

REFERENCES

1. *Energy Efficient Building Construction in Florida*, Florida Energy Extension Service, University of Florida, 2005. Adapted with permission, Gainesville, FL.
2. Jeffrey, T.; Dennis, C. *A Builder's Guide to Energy Efficient Homes in Georgia*; Georgia Environmental Facilities Authority; Division of Energy Resources: Atlanta, GA, 1999.
3. Environmental Building News. 2005. *Insulation Materials—Summary of Environmental and Health Considerations*. 122 Birge Street, Suite 30, Brattleboro, VT 05301. Phone: (800) 861-0954. Web site: http://www.BuildingGreen.com.
4. Federal Trade Commission. *Labeling and Advertising of Home Insulation (R-Value Rule) 16CFR460*. Available online at: http://www.ftc.gov/bcp/rulemaking/rvalue/16cfr460.htm Federal Register Notice (amendments) http://www3.ftc.gov/os/2005/05/05031homeinsulationfrn.pdf.
5. ToolBase Services. *Insulating Concrete Forms*. http://www.toolbase.org/techinv/techDetails.aspx?technologyID=97.
6. Oak Ridge National Laboratory, *Wall Insulation*. October 2000. Available online at: http://www.eere.energy.gov/buildings/info/documents/pdfs/26451.pdf.
7. Southface Energy Institute. March 2002. *Wall Insulation*. Available online at: http://www.southface.org/web/resources&services/publications/factsheets/26_insulatewalls4PDF.pdf.
8. Knowles, H. 2005. *Performance Under Pressure: Structural Insulated Panel (SIP) Walls*. http://www.energy.ufl.edu/factsheets.htm.
9. Desjarlais, A.O.; Petrie, T.W.; Stovall, T. In *Comparison of Cathedralized Attics to Conventional Attics: Where and When do Cathedralized Attics Save Energy and Operating Costs?*. Performance of Exterior Envelopes of Whole Buildings, IX International Conference ASHRAE: Clearwater, FL, 2004.
10. Oak Ridge National Laboratory, *Ceilings and Attics*. February 2000. Available online at: http://eber.ed.ornl.gov/DOEceilingsattics_771.pdf.
11. U.S. Department of Energy. *Radiant Barrier Attic Fact Sheet*. Available online at: http://www.ornl.gov/sci/roofs+walls/radiant/rb_05.html.

BIBLIOGRAPHY

1. Air Conditioning Contractors of America. Appendix A-5, detailed infiltration estimate. In *Manual J, Load Calculation for Residential Winter and Summer Air Conditioning*, 8th Ed.; May be ordered online at: http://www.accaconference.com/Merchant2/merchant.mv?ScreenZPROD&Store_CodeZACCOA&Product_CodeZ33-8&Category_CodeZ
2. Jeffrey, T.; Dennis, C. *Builder's Guide to Energy Efficient Homes in Louisiana*; Louisiana Department of Natural Resources: Louisiana, 2002.
3. Insulation Contractors Association of America. *Technical Bulletin No. 18, Reflective Insulations and Radiant Barriers*. Alexandria, VA, July 1997.
4. Oak Ridge National Laboratory, *Attic Access*. February 2000. Available online at: http://eber.ed.ornl.gov/Atticaccess_DOE_GO10099-768.pdf.
5. Office of Energy Efficiency and Renewable Energy (EERE) 1-877-337-3463 (EERE-INF) www.eere.energy.gov and http://www.eere.energy.gov/consumer/
6. U.S. Department of Energy. *Insulation and Weatherization*. Available online at: http://www.pueblo.gsa.gov/cic_text/housing/energy-savers/insulation.html.
7. U.S. Department of Energy. *Insulation Fact Sheet*. Available online at: http://www.ornl.gov/roofsCwalls/insulation.
8. U.S. Department of Housing and Urban Development, *Volume 2—The Rehab Guide: Exterior Walls*. Washington, DC, August 1999. HUD-DU100C000005956 Available online at: http://www.huduser.org/publications/destech/walls.html.
9. U.S. Department of Housing and Urban Development, *Volume 3—The Rehab Guide: Roofs*. Washington, DC, March 1999. HUD-DU100C000005956, Available online at: http://www.huduser.org/publications/destech/roofs.html.

Integrated Gasification Combined Cycle (IGCC): Coal- and Biomass-Based

Ashok D. Rao
Advanced Power and Energy Program, University of California, Irvine, California, U.S.A.

Abstract
Gasification or partial oxidation in an integrated gasification combined cycle (IGCC) consists of converting an often "dirty" fuel, such as coal, biomass, or refinery residues that cannot be directly used in gas turbines, to a clean gaseous fuel that meets engine specifications and environmental emissions standards. Gasification itself has been in commercial use for more than fifty years; its first applications involved making "town gas" for heating and cooking purposes before large natural gas reserves were discovered and providing synthesis gas (or syngas) for production of chemicals. Valuable coproducts such as H_2 or Fischer Tropsch liquids can enhance IGCC economics.

INTRODUCTION

Conversion of gasification feedstock is accomplished in a "gasifier," where the feedstock is typically reacted with an oxidant (oxygen or air) under partially oxidizing conditions. The partial oxidation reaction for carbon is

$$C + 1/2\,O_2 = CO.$$

Because this reaction is exothermic, water is introduced as steam or a liquid to moderate reaction temperature by the endothermic reaction:

$$C + H_2O = H_2 + CO.$$

Other reactions that occur within a gasifier include the shift reaction

$$CO + H_2O = H_2 + CO_2$$

and the hydrogasification reaction

$$C + 2H_2 = CH_4.$$

Sulfur present in feed is evolved mostly as H_2S with some as COS. CS_2 and mercaptans are insignificant when gasification occurs at high temperatures (over 1800°F or 980°C).

Prior to gasification reactions, feedstock undergoes heat-up, drying and pyrolysis or devolatilization. Pyrolysis produces a number of components such as H_2, NH_3, HCN, H_2S, COS, CS_2, mercaptans, CO_2, H_2O, CH_4, and higher hydrocarbons, tars and oils. Depending on the temperature of the gases leaving the gasifier, decomposition or reforming of some of the pyrolysis products may occur.

Typical syngas compositions obtained by high temperature gasification are shown in Table 1.

Keywords: Gasification zero emission; Gasifier; IGCC; Integrated gasification combined cycle.

TYPES OF GASIFIERS

There are three major types of gasifiers:

- Moving bed gasifiers as exemplified by the Lurgi gasifier,[15] consists of introducing coarse solids at the gasifier top while oxidant and steam are introduced at the gasifier bottom. The gasifier may be divided into four distinct zones, from top to bottom: (1) the drying/preheating zone, (2) the devolatilization zone, (3) the gasification zone, and (4) the combustion zone. Gases leave the top at relatively low temperatures (under 1000°F or 540°C) and contain H_2, CO, CO_2, H_2O, CH_4, and other organic compounds including oils and tars; sulfur compounds such as H_2S, COS, some CS_2 and mercaptans; and nitrogen compounds such as NH_3 and HCN. The tars also contain some sulfur and nitrogen. Gasifier thermal efficiency is high while specific O_2 consumption is low, but since the process produces tars and oils, gas clean-up is more complex. Only a limited amount of fines in the feed can be handled. Coals with high Free Swelling Indices could plug up the bed and are not desirable feedstocks. The British Gas Lurgi gasifier,[3] which is similar to the original Lurgi in many respects, recycles tars and oils separated from the gas to the gasifier by introducing these into the bottom section of the gasifier along with fines.
- The fluidized bed gasifier as exemplified by the Advanced Transport Reactor[8] consists of introducing dried feed that is typically less than 0.25 in. (0.64 cm) near the bottom of the gasifier and accomplishing devolatilization and gasification in a bed fluidized by oxidant and steam. Recycled synthesis gas may also be used to maintain required gas velocity for fluidization of bed material. In the bottom section of the gasifier, the temperature is high enough to fuse ash

Table 1 Typical syngas compositions

Component (vol.% dry basis)	Air blown Biomass feed	High temperature entrained Oxygen blown High sulfur coal feed
CO	15	46.5
H_2	16	37.3
CO_2	18	13.5
CH_4	7	0.1
N_2+Ar	44	1.9
H_2S+COS	Trace	0.7
Total	100	100.0

particles together to form agglomerates; while in the top of the gasifier, where gasification reactions predominate, temperatures are lower (1800°F–1900°F or 980°C–1040°C). Gases leaving the gasifier are essentially free of hydrocarbons heavier than CH_4. Thermal efficiency of this type of gasifier tends to be lower since syngas carries more of feed (coal or biomass) bound energy as sensible heat than moving bed gasifiers, while specific O_2 consumption tends to be higher. Carbon conversion tends to be rather low with unreactive feedstocks, such as high-rank coals, and is more suitable to lower-rank coals (lignites and subbituminous) and biomass. This gasifier is especially suitable for biomass applications in which the feedstock also tends to be heterogeneous with respect to particle size and composition. Further, since a large inventory of bed material is maintained within the gasifier, the gasifier performance tends to be less susceptible to sudden changes in feedstock composition.

- Entrained bed gasifiers, as exemplified by General Electric[9] and Shell,[5] operate at temperatures typically in excess of 2200°F or 1200°C, with ash forming a slag, while syngas is essentially free of hydrocarbons heavier than CH_4. In the General Electric gasifier, solids are introduced in the form of water slurry (containing finely ground solids); while in the Shell gasifier, solid feed is fed, after drying and grinding, with steam used as a temperature moderator. Thermal efficiency for a given carbon conversion tends to be the lowest among the three types of gasifiers, since syngas carries a significant portion of feed-bound energy as sensible heat, while specific O_2 consumption tends to be the highest. This type of gasifier is more suitable to unreactive feedstocks, such as high-rank coals and petroleum coke. The E-Gas entrained bed gasifier[6] introduces coal as a slurry into two sections of the gasifier in order to increase its efficiency while decreasing the specific O_2 consumption.

Depending on the type of gasifier, feedstock, and end use for the syngas, the operating pressure may vary from near atmospheric to as high as 1200 psig (8400 kPag).

Novel types of gasifiers are being developed, such as those that react steam with feedstock while using heat provided by an inert solid-circulating medium heated externally.[10] Such gasifiers have the advantage of utilizing air in the process instead of oxygen; the circulating medium is heated externally by combusting the unreacted char with the air while producing syngas with a heating value similar to or greater than that obtained from Oxygen blown gasifiers.

FEEDSTOCKS

Feedstocks that may be converted to syngas by gasification:

- Coal
- Biomass
  - — Agricultural
  - — Municipal (Sewage Sludge, Solid Waste)
- Paper Mill Black Liquor
- Refinery Residues
  - — Petroleum Coke
  - — Asphaltenes, Visbreaker Bottoms
  - — Refinery and Petrochemical Wastes
- Oil Emulsion (Orimulsion)

Important feedstock characteristics to be taken into account for gasifier selection (Also dependent on ultimate use of syngas):

- Reactivity
  - — Limited carbon conversion for fluidized bed gasifiers with lower reactivity feedstocks.
- Moisture Content
  - — High moisture content a thermal penalty for dry pulverized feed gasifiers, as well as for slurry feed gasifiers when it is bound moisture.
- Ash Content
  - — High ash content a thermal penalty for slurry feed gasifiers.
  - — Low ash content a challenge for membrane wall gasifiers such as the Shell gasifier; these may require recycle of ash.
- Ash Fusion Temperature Under Reducing Conditions (typically lower than that under oxidizing conditions for coals, while higher for petroleum cokes)
  - — High ash fusion temperature a thermal penalty for slagging gasifiers while shortening refractory life for refractory lined gasifiers; may require fluxing agent.
  - — Low ash fusion temperature a challenge for non-slagging gasifiers; may increase gasifier steam consumption.

- Particle Size
  — Lower limit on fines fraction in feed to moving bed gasifiers; may require briquetting.
- Free Swelling Index
  — Can cause plugging in moving bed gasifiers; may require a stirrer.

SYNGAS TREATMENT

Raw gas leaving a gasifier is either cooled by heat exchangers while generating steam or directly quenched with water. The cooled gas is purified by further treatments to remove contaminants such as particulates, alkalis, chlorides, and nitrogen and sulfur compounds. Various contaminants that could be present in coal-derived raw syngas are listed in Table 2. The exact nature of contaminants and their concentrations is dependent on the type of feedstock and type of gasifier. For example, potassium salts can be in significant concentrations in the case of raw syngas derived by the gasification of agricultural products, while the chlorides can be in significant concentrations in municipal-solid-waste-derived syngas. An advantage with gasification is that sulfur may be recovered as salable elemental sulfur or sulfuric acid.

The following section describes "cold gas cleanup" technologies used for treatment of syngas before it can be fired in gas turbines. "Warm gas cleanup" technologies to clean gas at higher temperatures (570°F–750°F or 300°C–400°C) are being developed, and some of these are in the demonstration phase.[2,4,7] Syngas is kept above its water dew point, which eliminates extensive waste water treatment typically required by gasification using cold gas cleanup.

Acid Gas Removal

Acid gas removal processes available for capturing sulfur compounds from syngas include the following:

1. Selective Amine Scrubbing (aqueous methyl diethanol amine solvent with additives)
2. Rectisol (methanol solvent)
3. Selexol® (mixture of dimethyl ethers of polyethylene glycol solvent)

The amine solvent, being a chemical solvent, requires a significant amount of steam for its regeneration to release absorbed acid gases. Additives improve selectivity between H_2S and CO_2 absorption and are suitable for applications in which CO_2 capture is not required. Without good selectivity, a significant fraction of CO_2 is co-absorbed by the solvent. The loss of CO_2 from syngas has the detrimental effect of reducing the amount of motive fluid (available at pressure) for expansion in gas turbines while increasing the amount of diluent required for NOx control. Furthermore, solvent circulation rate and utility consumptions are increased. The efficiency and cost of downstream sulfur recovery units (SRUs) (for converting captured H_2S and any co-absorbed COS to elemental sulfur), and tail gas treating units (for hydrogenating trace amounts of elemental sulfur and its oxygen containing compounds and for recovery of the H_2S) are also negatively affected due to the presence of a large concentration of CO_2 in the acid gas.

The amine process is used both in the Polk County, Florida[9] and Wabash, Indiana[6] IGCC plants. In Wabash, tail gas is recycled to the gasifier after hydrogenation, instead of treating it in an amine unit to recover the H_2S for recycling to the SRU.

Both the amine and Selexol® processes require COS hydrolysis to convert COS to H_2S, since only a small fraction of COS is captured. Hydrolysis is accomplished by the catalytic reaction:

$$COS + H_2O = H_2S + CO_2.$$

Rectisol and Selexol® processes utilize physical solvents and work best when the syngas to be treated is at a high pressure, typically greater than 400 psi (2800 kPa) for Rectisol and greater than 600 psi (4100 kPa) for Selexol®. Refrigeration is required to cool the solvent to enhance solubility.

Rectisol is especially suitable when very low specification on syngas sulfur content is required. COS hydrolysis is not required since it is readily absorbed by the solvent. Selexol® is especially suitable in "Zero Emission" IGCC plants where CO_2 emissions are to be controlled.[11] Syngas is decarbonized and desulfurized in the same Selexol® unit to produce gas that is essentially H_2 for combustion in gas turbines. CO_2 is released from the solvent at pressures significantly higher than atmospheric to reduce the power consumption of compressors required to generate a high-pressure CO_2 stream for sequestration.

Table 2 Typical coal derived syngas contaminants

Particulates	Carbonyls	Sulfur compounds
Ash	Ni	H_2S
Soot	Fe	COS
		CS_2
		Mercaptans
Volatile metals	**Chlorides**	**Nitrogen compounds**
Hg	NaCl	NH_3
As	HCl	HCN
Cd		
Se		

Indu–Inter

Metal Carbonyls

Metal carbonyls may be captured by activated carbon. Carbon particles accompanying syngas, leaving the gasifier, and collecting in barrier filters located upstream of the raw syngas scrubber may also capture carbonyls. Rectisol solvent captures carbonyls by converting them into sulfides.

Mercury, Arsenic, Cadmium and Selenium

Mercury, arsenic, cadmium and selenium typically volatilize within the gasifier and leave with the raw syngas. Sulfided activated carbon has been used to remove mercury and arsenic from syngas at the Tennessee Eastman plant.[14] Calgon offers this activated carbon for removal of mercury, reducing its concentration to as low as 0.01–0.1 μg/Nm^3 Hg in syngas depending on operating temperature and moisture content. Mercury is captured predominantly as a sulfide, with some captured as elemental mercury. Its capture is reduced as syngas temperature and relative humidity are increased. Spent carbon is disposed of as hazardous waste, although attempts are being made to recover elemental mercury.

Experience at the Eastman plant indicates that activated carbon is more effective in capturing arsenic. Calgon's experience has shown that arsenic, if present as an arsine, is captured by this sulfided carbon. SudChemie offers activated carbons for removal of arsenic and its compounds. Copper-impregnated carbon is offered to capture arsenic if present as an organic compound. The capture of other volatile metal (cadmium and selenium) compounds by the activated carbon is also expected to occur.

AIR SEPARATION TECHNOLOGY

In high-temperature gasifiers, gasifier efficiency is increased and the size of downstream equipment is minimized when oxygen is used instead of air. If air is used, a significant portion of feed-bound "chemical energy" is degraded to thermal to heat the N_2 to gasifier effluent temperature. In plants where CO_2 capture is required or syngas is to be utilized for co-producing products such as H_2 or methanol, it is essential that the gasifier be oxygen blown. The use of air as an oxidant in a gasifier may be more attractive in certain cases:

- Smaller-scale plants (e.g., biomass applications) where the cost of producing O_2 at smaller scales becomes unattractive.
- Lower-temperature gasifiers such as fluidized beds, especially if limestone is utilized to capture significant portion of sulfur within the gasifier itself.

Cryogenic Technology

Conventional air separation technology for large-scale O_2 production consists of cryogenically separating air. For IGCC applications, an Elevated-Pressure (EP) Air Separation Unit (ASU) is preferred over Low-Pressure (LP) ASUs, because both O_2 (for gasifier) and N_2 (for NOx control and increased power output in gas turbine) can be utilized at elevated pressures. The distillation pressure within the ASU cold box affects the liquid bubble-point. As this pressure is increased, cold-box temperature increases, resulting in a reduced ratio of incoming air pressure to outgoing stream pressures (O_2 and N_2). Thus, if O_2 and N_2 can be utilized within the plant at the cold-box outlet pressure or higher, a net increase in the IGCC plant efficiency is realized. N_2 produced by an EP cold box is further compressed and fed to gas turbines.

The relative volatility between N_2 and O_2 approaches unity as this operating pressure is increased, requiring more stages in the distillation operation; thus, an optimum upper limit exists for the pressure.

Typically, a 2% reduction in plant heat rate and cost may be realized with EP ASU over LP ASU. Optimum O_2 purity for IGCC applications with cryogenic ASUs (with either EP or LP) is 95%. The number of distillation stages decreases steeply as purity is reduced from 99.5 to 95%, but remains quite insensitive as purity is further reduced. O_2 compression costs (both capital and operating) continue to increase as purity is decreased below 95%. The size of most equipment downstream of the ASU also increases slightly while efficiency of the gasification unit decreases as purity is reduced. Both the Demkolec[5] and Polk County IGCCs utilize an EP ASU (with 95% purity O_2). Higher-purity O_2 may be required in coproduction plants (e.g., coproducing H_2 utilizing pressure swing adsorption to limit its Ar content).

Feed air pressure for an LP ASU is in the range of 50–90 psig (350–600 kPag), while that for an EP ASU is typically set by the pressure of air extracted from gas turbines (to provide the portion of air required by the ASU). Extraction of air from the gas turbine compressor discharge increases commonality for the gas turbine (relative compressor/turbine size) between IGCC and natural gas applications. When extraction air pressure is very high, partial expansion may be required in order to limit the cold-box operating pressure so the relative volatility between O_2 and N_2 is not too close to unity.

High-Temperature Membrane Technology

Cryogenic ASU power consumption constitutes more than half of the total power consumed by the IGCC or 5%–10% of total power produced. Membranes (semi-conductor materials) are being developed that operate at temperatures around 1500°F–1600°F (800°C–900°C) for air separation with a goal of reducing power consumption and capital cost by about 30%. Integration of the membrane unit with a gas turbine capable of roughly 50% air extraction (not possible with current large-scale machines) is required for this technology to be economical, with the feed air to the membrane being the extracted high-pressure air (preheated by directly firing syngas). Gas turbines also capable of receiving depleted air from the membrane unit are required. A large-scale membrane unit with a capacity of 2,000 ST/D (1800 MT/D) is expected to be available for demonstration in 2012.[1]

OVERALL PLANT INTEGRATION

Fig. 1 depicts a coal-based IGCC consisting of slurry-fed, high-pressure entrained bed gasifiers. Cryogenic EP ASU supplies 95% purity O_2 to gasifiers and produces intermediate pressure N_2 for injection into gas turbines. Coal is wet ground in rod mills to form slurry and introduced into the gasifiers. The gasifiers generate hot gas (raw syngas), slag, and char particulates. Raw syngas is cooled while generating high-pressure steam in heat exchangers (radiant coolers followed by convective coolers). The gas is then wet scrubbed to remove entrained solids. Soluble alkali salts, hydrogen halides, and some NH_3 are also removed. The contaminated water is treated to remove fine slag and char particulates, and the remaining water is treated before discharge.

Scrubbed gas is preheated and then fed to a COS hydrolysis reactor. Effluent then enters a series of heat exchangers to generate steam, heat up circulating water for a syngas humidifier (to introduce water vapor into clean syngas for NOx control and increase motive fluid in the gas turbine), and vacuum condensate from the combined cycle unit. After cooling against cooling water, the syngas is superheated by about 20°F (11°C) to avoid condensation and then fed to activated carbon bed(s) to remove volatile metals and their compounds. The effluent gas is then cooled against cooling water and fed to an Acid Gas Removal Unit (AGRU). The high-temperature condensate separated from the gas is recycled to the scrubber while the colder condensate is fed to a sour water stripper to remove absorbed gases such as NH_3, HCN, and H_2S. The sour gases stripped from the water are routed to a Claus SRU.

The acid gas removal process consists of an absorber column where H_2S is removed. Rich solvent loaded with H_2S after preheating (against stripped solvent) is fed to a stripper to release a concentrated stream of H_2S along with co-absorbed CO_2. Stripping is accomplished using steam. After the heat exchange, stripped solvent is recycled back to the absorber. Acid gas mixture from the stripper is cooled, and the condensate is removed to be sent to the SRU (to produce elemental sulfur) and a Tail Gas Treating Unit.

Clean (treated) syngas, after humidification in a counter-current column (by direct contact with hot water) and preheating, is fed to gas turbines. Exhaust gas

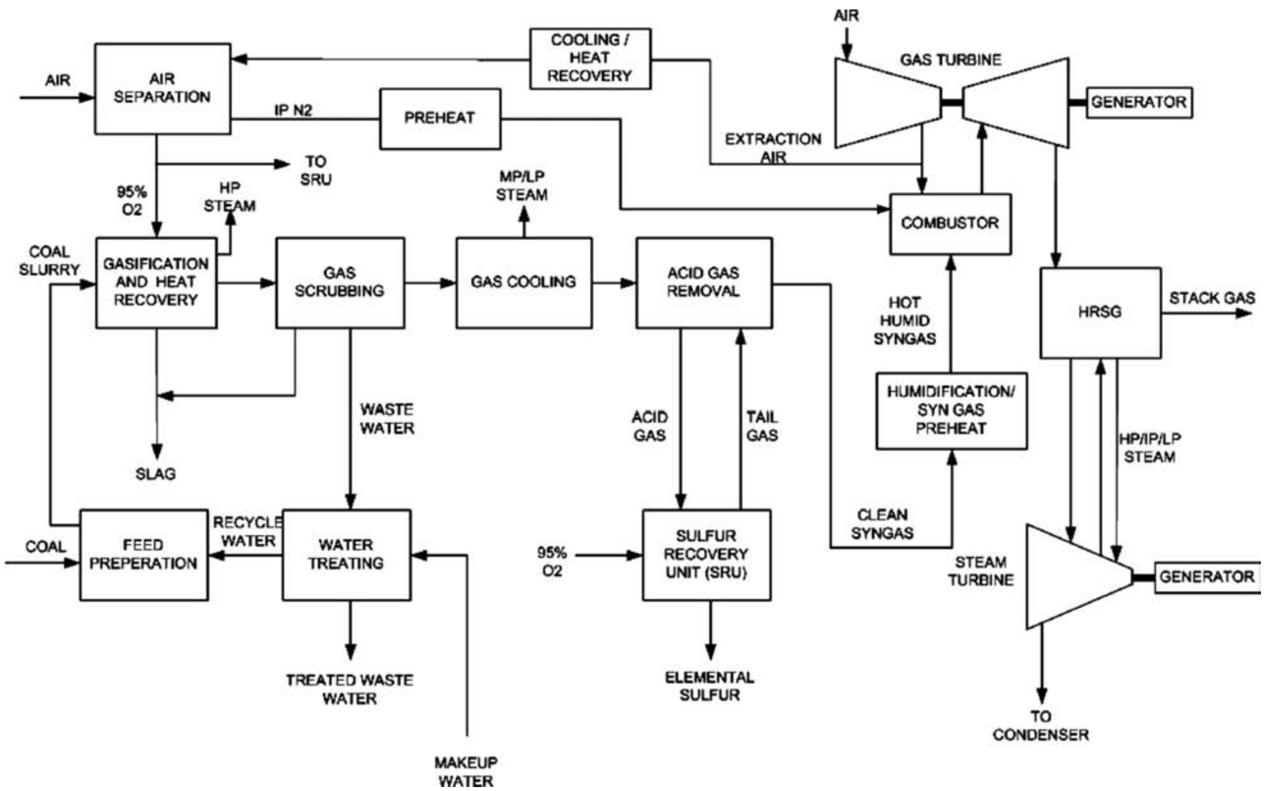


Fig. 1 IGCC overall block flow sketch.

from the gas turbines is fed to triple-pressure Heat Recovery Steam Generators (HRSGs), which provide superheated and reheated steam to a condensing steam turbine.

Necessary general facilities required include cooling water systems, instrument air, and flare.

Coproduction

A major advantage of an IGCC is that a number of co-products may be produced to improve process economics. Fig. 2 depicts various products and co-products that may be produced. Advantages include the following:

- Economies of scale of larger plants
- Efficient use of steam in a single large steam turbine
- Opportunity for load following
- Advantages specific to products such as methanol or Fischer Tropsch liquids:
  - — Reduced synloop recycle rate with unconverted purge gas fired in gas turbines
  - — Higher reactor throughput due to lessened buildup of inerts
  - — Reduction in diluent addition to gas turbines for NOx control due to low heating value of purge gas

Environmental Signature

Integrated gasification combined cycle is one of the most efficient, environmentally effective means of producing electricity. Air emissions from an IGCC are far below U.S. Clean Air Act standards. Reductions of emissions of SO_x, NOx, CO, volatile metals, and particulates can be significantly better than those obtained by scrubber-equipped pulverized, as well as circulating fluidized bed coal combustion plants.

Integrated gasification combined cycle's economic benefits will become more significant as air emission standards are made more stringent, since this technology can achieve greater emissions reductions at lower incremental costs.

Integrated gasification combined cycle plants can be designed to approach zero emissions, including CO_2 emissions.[12] Carbon is captured as CO_2 from syngas, such that a gas that is mostly H_2 is burnt in gas turbines. This pre-combustion capture may be combined with syngas desulfurization and may be accomplished in a physical solvent acid gas removal process (e.g., Selexol®). Major modifications to the process scheme described in the previous section include the following:

- Adding catalytic shift reactors to convert CO in syngas to CO_2 by the reaction

$$CO + H_2O = CO_2 + H_2.$$

- Use of a multi-column design in the AGRU to capture CO_2.

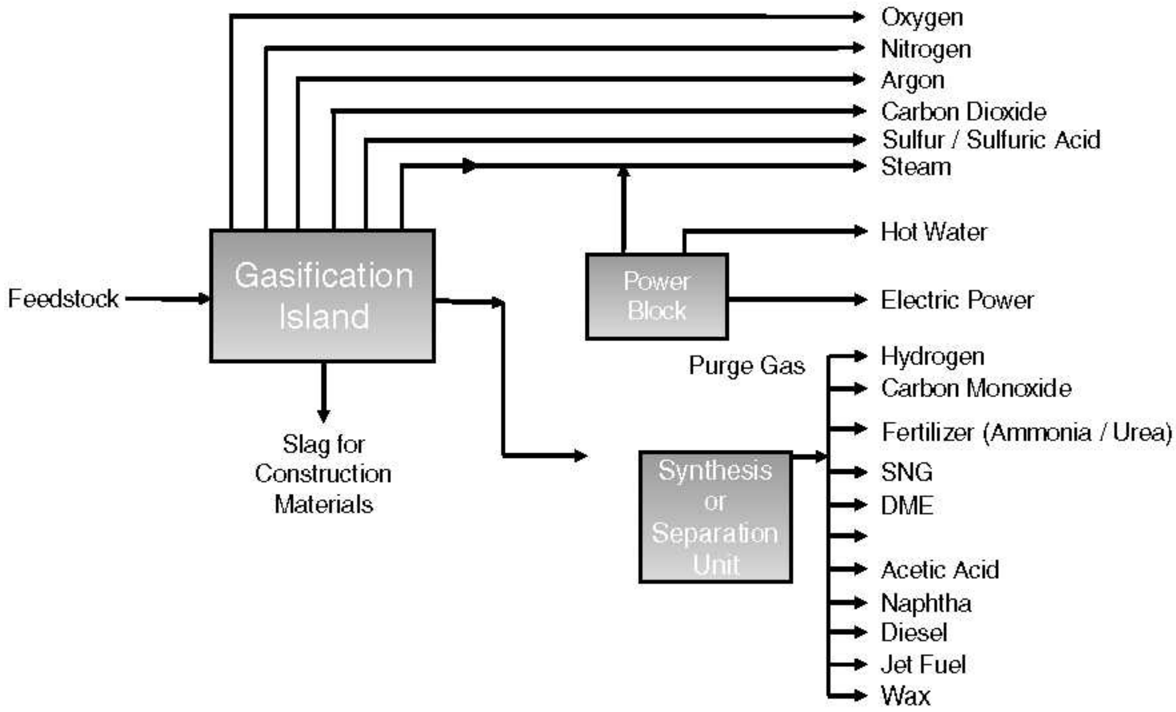


Fig. 2 Coproduction alternatives in an IGCC.

IGCC VERSUS PULVERIZED-COAL BOILER FUTURE TRENDS

In the case of high-rank coals, economics are typically in favor of an IGCC when compared to a supercritical coal-fired boiler plant, with environmental constraints very stringent. The cost of 90% or greater volatile mercury removal from a coal IGCC is about a tenth of that from a pulverized coal combustion plant.[13] In addition to lower cost, gasification-based mercury removal is more efficient than removal of mercury after combustion because the volume of pressurized syngas is much smaller than stack gas from the pulverized coal plant, which is at atmospheric pressure and diluted with N_2.

Technological advancements aimed at increasing plant thermal efficiency of both types of plants are taking place. Advanced gas turbines are being developed that will have a significant effect on both IGCC thermal efficiency and plant cost (Fig. 3). As power block efficiency increases, the amount of syngas required to generate a unit of electric power is reduced in direct proportion, which in turn reduces plant size and cost upstream of the power block.

ECONOMIES OF SCALE AND BIOMASS GASIFICATION

Plant size has a significant effect on IGCC economics. The typical size required for a commercial plant to take advantage of economies of scale is in excess of 400 MW, while smaller decentralized plants fit with the smaller biomass resources. Due to its low energy density, transportation of biomass collected from various locations over long distances to a central plant is economically prohibitive. The biomass facility has to be located in close proximity to the feedstock. Consideration should be given to building a larger IGCC fed by both biomass and coal (or petroleum coke). During seasons of low biomass feedstock availability, the plant would increase its coal usage to offset a decrease in biomass-derived syngas. In the case of co-feeding an IGCC with a high-rank coal or petroleum coke with biomass, two types of gasifiers may be required, one optimized for the coal or petroleum coke such as an entrained bed gasifier, and the other suitable for the biomass feedstock such as a fluidized bed gasifier.

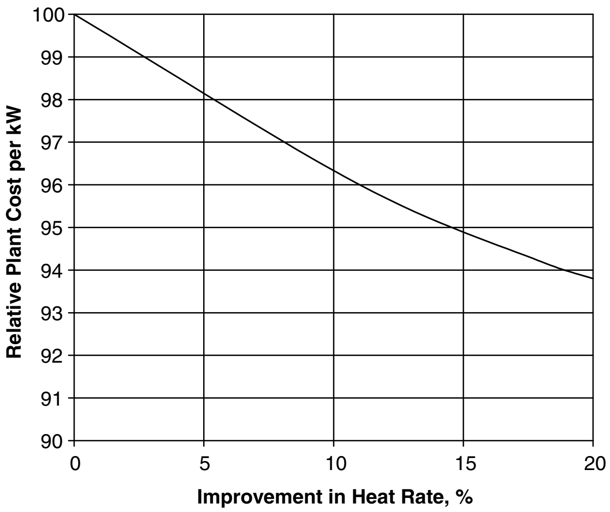


Fig. 3 Effect of power block efficiency on IGCC cost.

CONCLUSIONS

Integrated gasification combined cycle is commercially proven, and the gasification process itself has been in commercial use for more than 50 years. Valuable coproducts such as H_2 or Fischer Tropsch liquids may be co-produced to enhance economics. Integrated gasification combined cycle is one of the most efficient, environmentally effective means of producing electricity. Integrated gasification combined cycle plants can be designed to approach zero emissions.

REFERENCES

1. Armstrong, P.A. ITM oxygen: the new oxygen supply for the new IGCC market. Presented at the Gasification Technologies Conference, San Francisco, October, 2005.
2. Butz, J.R.; Lovell, J.S.; Broderick, T.E.; Sidwell, R.W.; Turchi, C.S.; Kuhn, A.K. Evaluation of amended silicates™ sorbents for mercury control. Proceedings of the Mega Symposium, Washington, DC, May 2003.
3. EPRI Report No. GS-7091. U.S. Bituminous coal test program in the british gas/lurgi (BGL) gasifier, December 1991.
4. Gangwal, S.K.; Turk, B.S.; Coker, D. et al. Warm-gas desulfurization process for chevron texaco quench gasifier syngas. Presented at the 21st Annual Pittsburgh Coal Conference, Pittsburgh, Pennsylvania, 2004.
5. Hannemann, F.; Kanaar, M. Operational experience of the buggenum IGCC plant & enhanced V94.2 gas turbine concepts for high syngas reliability and flexibility. Presented at the Gasification Technologies Conference, San Francisco, CA, October 2002.
6. Keeler, C.G. Wabash river in its fourth year of commercial operation. Presented at the Gasification Technologies Conference, San Francisco, CA, October 1999.
7. Krishnan, G.N.; Gupta R.; Ayala, R.E. Development of Disposable sorbents for chloride removal from high-temperature coal-derived gases. DOE/MC/30005-97/C0748; CONF-960757-38, 1996.
8. Leonard, R.; Rogers, L.; Vimalchand, P.; Liu, G.; Smith, P.; Longanbach, J. Development status of the transport gasifier at the PSDF. Presented at the Gasification Technologies Conference, San Francisco, CA, October 7–10, 2001.
9. McDaniel, J.E.; Shelnut, C.A. Tampa electric company polk power station IGCC project—Project status. Proceedings

Indu–Inter

of the Gasification Technologies Conference, San Francisco, CA, October 1999.
10. Paisley, M.A.; Irving, J.M.; Overend, R.P. A promising power option—The ferco silvagas biomass gasification process—operating experience at the burlington gasifier. Proceedings of the ASME TURBO EXPO, New Orleans, Louisiana, June 4–7, 2001.
11. Rao, A.D.; Verma, A.; Samuelsen, G.S. Acid gas removal options for the futuregen plant. Proceedings of the Clearwater Coal Conference, Clearwater, FL, April 2005.
12. Rao, A.D.; Stobbs, R. An evaluation of coal gasification with CO2 capture. Proceedings of the Combustion Canada Conference, Vancouver, September 2003.
13. Stiegel, G.S.; Longanbach, J.R.; Klett, M.G.; Maxwell, R.C.; Rutkowski, M.D. The cost of mercury removal in an IGCC plant. DOE/NETL Final Report, September 2002.
14. Trapp, W.; Denton, D.; Moock, N. Gasification plant design and operation considerations. Presented at the Gasification Technologies Conference, San Francisco, CA, October 2002.
15. van Dyk, J.C.; Keyser M.J.; Coertzen, M. Sasol's unique position in syngas production from South African coal sources using sasol lurgi fixed bed dry bottom gasifiers. Presented at the Gasification Technologies Council Conference, Washington, DC, 2004.

IntelliGrid^{SM}

Clark W. Gellings
Electric Power Research Institute (EPRI), Palo Alto, California, U.S.A.
Kurt E. Yeager
Galvin Electricity Initiative, Palo Alto, California, U.S.A.

Abstract
The IntelliGrid (IntelliGrid is a service mark of the Electric Power Research Institute.) is a fully functional power delivery system that incorporates sensors, communications, enhanced computational ability, and electronic control. It delivers reliable, digital-grade power to meet the needs of an increasingly electronic society.

The IntelliGrid is a possible foundation for the intelligent electric infrastructure of the future and the platform for future technical, process, and service innovation in the energy sector.

INTRODUCTION

Today's electricity supply system is aging, under stress, often not well maintained, and being physically used in ways for which it was not designed.

Society has entered a new era of experience driven by digital technologies. The world is more interconnected than at any other time in history, utterly dependent on the integrity of a complex web of networks—both communications and electricity. The electric power system is the foundation of this interconnection. The lack of critical infrastructure investment and surging demand for electricity are taxing the electric power system to its limits. In the past ten years, growth in U.S. electricity demand has exceeded new transmission capacity by more than 15%. This occurs at a time when microprocessor-based technologies have radically altered the nature of the demand for electricity, resulting in demand that is compatible with a power system created to meet the needs of an analog economy. This has led to quality and reliability problems responsible for more than $100 billion in losses to U.S. industry and society annually.

A coordinated effort to upgrade and modernize the entire electric power system—from generation to transmission, distribution, and end use—is needed for the benefits or interconnection to be fully realized.

DESIGNING THE POWER SYSTEM OF THE FUTURE

The IntelliGrid Consortium is a broad-based collaboration of energy, high-tech, and government leaders working to address these technology issues and set us on a migration path toward the intelligent, self-healing power system of the future.

The foundation of this system is the IntelliGrid—an open systems-based comprehensive reference architecture for the energy enterprise of the future. IntelliGrid enables the integration of intelligent equipment and data communications networks into a managed enterprise and industrywide distributed computing system. It is the basis for enabling enhanced system capabilities, such as the self-healing grid, integrated consumer communications, and real-time energy information and power exchanges.

IntelliGrid applies the latest systems engineering methods and computing and communications tools to model the energy enterprise of the future. It expands traditional operating boundaries, promoting greater interoperability; enabling unprecedented improvement in performance and asset utilization; and ensuring desirable levels of robustness, responsiveness, and security.

The IntelliGrid is the basis of a "smart" power delivery system. It is an integrated, self-healing, self-diagnosing, electronically controlled electricity supply system of extreme resiliency and responsiveness, one that is fully capable of responding in real time to consumer needs.

The IntelliGrid will be an integrated electricity and information infrastructure that enables the next wave of technological advances to flourish. The system will be always on and "alive," interconnected and interactive, and merged with communications in a real-time network, sensing and communicating information and power exchange. The IntelliGrid will be a self-healing system that senses disturbances and corrects them, or reconfigures the flow of power to mitigate interruption and disturbance. The system will also be smart enough to integrate traditional power generation with distributed energy resources (DER) seamlessly.

Keywords: IntelliGrid; Smart power delivery system; Self-healing; Integrated consumer communications; Transmission; Distribution; Consumer portal; FACTS; Communications.

The goal of the IntelliGrid is to create a new self-healing paradigm for power delivery systems, with automated capabilities that can anticipate many potential problems, reduce recovery time when unexpected disturbances occur, and enhance performance of normal operations.

To reach this goal, three primary objectives of the self-healing grid need to be achieved:

- Dynamically optimize the performance and robustness of the system
- Quickly react to disturbances in such a way as to minimize impact
- Effectively restore the system to a stable operating region after a disturbance

The initial step in developing the IntelliGrid is to define clearly the scope of the requirements of the power system functions and to identify all the roles of the stakeholders. These then must be included in an Integrated Energy and Communications System Architecture. Many power system applications and a large number of potential stakeholders already participate in power system operations. In the future, more stakeholders—such as customers responding to real-time processes, distributed energy resource owners selling energy and ancillary services into the electricity marketplace, and consumers demanding high-quality power—will participate actively in power system operations. At the same time, new and expanded applications will be needed to respond to the increased pressures for managing power system reliability as market forces push the system to its limits.

Power system security has also been recognized as crucial in the increasingly digital economy. The key is to identify and categorize all these elements so that their requirements can be understood; their information needs can be identified; and eventually, synergies among these information needs can be determined. One of the most powerful methodologies for identifying and organizing the pieces in this puzzle is to develop business models that identify a straw set of entities and address the key interactions among these entities. These business models will establish a set of working relationships among industry entities in the present and the future, including intermediate steps from vertical operations to restructure operations. The business models will include, but not be limited to, the following:

- Market operations, including energy transactions, power system scheduling, congestion management, emergency power system management, metering, settlements, and auditing
- Transmission operations, including optimal operations under normal conditions, prevention of harmful contingencies, short-term operations planning, emergency control operations, transmission maintenance operations, and support of distribution system operations
- Distribution operations, including coordinate volt/var control automated distribution operation analysis, fault location/isolation, power restoration, feeder reconfiguration, distributed energy resource management, and outage scheduling and data maintenance
- Customer services, including automatic meter reading (AMR), time-of-use and real-time pricing, meter management, aggregation for market participation, power quality monitoring, outage management, and in-building services and services using communications with end-use loads within customer facilities
- Generation at the transmission level, including automatic generation control, generation maintenance scheduling, and coordination of wind farms
- Distributed resources (DR) at the distribution level, including participation of DR in market operations, DR monitoring and control by nonutility stakeholders, microgrid management and DR maintenance management

These business models will be analyzed and used to define the initial process boundaries for subsequent tasks. Business process diagrams will be used to illustrate the more complex interactions.

FAST SIMULATION AND MODELING

When the architecture begins to be deployed, the industry's computational capability must be enhanced. To enable the enhanced functionality of the power delivery system, a capability that allows fast simulation and modeling (FSM) will need to evolve to ensure the mathematical underpinning and look-ahead capability for a self-healing grid—one capable of automatically anticipating and responding to power system disturbances while continually optimizing its own performance. Creating a self-healing grid will require the judicious use of numerous intelligent sensors and communication devices that will be integrated with power system control through a new architecture. This new architecture will provide a framework for fundamentally changing system functionality as required in the future.

The FSM will augment these capabilities in three ways:

- Provide faster-than-real-time, look-ahead simulations and to avert unforeseen disturbances
- Perform what-if analysis for large-region power systems from both operations and planning points of view
- Integrate market, policy, and risk analysis into system models, and quantify their effects on system security and reliability

The next step in creating a self-healing grid will involve addition of Intelligent Network Agents (INAs) that gather and communicate system data, make decisions about local control functions (such as switching a protective relay), and coordinate such decisions with overall system requirements. Because most control agents on today's power systems are simply programmed to respond to disturbances in predetermined ways—by switching off a relay at a certain voltage, for example—their activity may actually make an incipient problem worse and contribute to a cascading outage.

The new simulation tools developed as part of EPRI's FMS project will help prevent such cascading effects by creating better system models that use real-time data coming from INAs over a wide area and, in turn, coordinate the control functions of the INAs for overall system benefit, instead of the benefit of one circuit or one device. As discussed later in this article, having such improved modeling capability will also enable planners to better determine the effects of various market designs or policy changes on power system reliability.

To reach these goals, the FSM project will focus on the following three areas:

- *Multiresolution modeling*. New modeling capabilities will be developed that provide much faster analysis of system conditions and offer operators the ability to "zoom" in or out to visualize parts of a system with lower or higher degrees of resolution.
- *Modeling of market and policy impacts on reliability*. The recent power crisis in the Western states dramatically illustrated how untested policies and market dynamics can affect power system reliability. The new modeling capabilities being developed in this project will allow planners and policymakers to simulate the effects of their activities before putting them into practice.
- *Validation of integrated models with real-time data*. Once the new, integrated models have been thoroughly tested offline, they will be validated and enhanced using real-time data from major power systems.

Unified integration architecture is the key enabler to deploying advanced functions successfully and inexpensively. This architecture must be robust enough to meet the numerous disparate requirements of power system operations and must be flexible enough to handle changing needs. The focus of this project is to identify and propose potential solutions for enterprise and industrywide architectural issues that will arise from the high levels of integration and management foreseen by this project.

The IntelliGrid will be constantly self-monitoring, communicating, and self-correcting to maintain the flow of secure, digital-grade quality, and highly reliable and available power.

The IntelliGrid concept includes a combination of enhancements to the power delivery infrastructure and consumer technologies to enable delivery of highly reliable digital-grade power. The IntelliGrid enhancements will result in very long mean time between failure (MTBF) rates, thereby maximizing reliability for digital devices, digital applications, and the digitally enabled enterprises they make possible.

The IntelliGrid architecture enables equipment manufacturers to design products to function at the reliability levels that they realistically encounter in the field. Manufacturers could embed resilience to degraded power quality and reliability into their products, provided that an industry consensus is reached and/or market forces motivate them to do so. The IntelliGrid is expected to be a key to economic prosperity at all levels of the economy.

The IntelliGrid architecture includes a window to the consumer referred to as the consumer portal or EnergyPort. The EnergyPort relieves the constraints provided by the meter, allowing price signals, decisions, communications, and network intelligence to flow back and forth through a two-way portal. The EnergyPort is the integration technology that leads to a fully functioning retail power marketplace with consumers responding to price signals. Capabilities of the EnergyPort can include the following:

- Real-time pricing
- Value-added services
- Building and appliance standards
- Consumer energy management
- Distributed energy resources (DER)
- Real-time system operations
- Short-term load forecasting
- Long-term planning

The IntelliGrid includes functional attributes that address the efficiency of electricity use. Advances in electrotechnologies, advanced motors, lighting, improved space conditioning, real-time energy management systems, automated recycling systems, and other technologies could encourage investment in energy-efficient technologies, reduce energy costs, stimulate the economy, and minimize the environmental impact of electricity generation and delivery. In addition, widespread use of electric transportation solutions can enhance energy balance.

The IntelliGrid will also encourage widespread deployment of DER.

Several critical enabling technologies will enable the IntelliGrid to evolve, including the following:

- Increased power flow
- Automation
- Communications architecture
- Distributed energy resources (DER)

- Power electronics-based controllers
- Power market tools
- Technology innovation
- The EnergyPort
- Value-added services

Increased Power Flow

Real and reactive power flows in large integrated power transmission systems are constrained by voltage and power stability criteria. In many cases, the limits dictated by these stability criteria are far below the "inherited" thermal capacity of transmission corridors. Even with the adoption of stability measures, power flow levels could still be below the thermal limit due to the "uncontrolled" flow levels as determined by the power transfer law, which is governed by the line impedance, voltage magnitudes, and angle difference at the transmission corridor ends.

This technological solution is named FACTS (Flexible AC Transmission Systems). The use of FACTS controllers helps preserve system integrity during the short-term major system disturbances or to direct and control real and reactive power flows fully on transmission corridors.

Automation: The Heart of the IntelliGrid

Automation will play a key role in providing high levels of reliability and quality. To a consumer, automation may mean receiving hourly electricity price signals, which can adjust home thermostat settings automatically. To a distribution system operator, automation may mean automatic "islanding" of a portion of the distribution system in an emergency. To a power system operator, automation means a self-healing, self-optimizing power delivery system. Through the EnergyPort, consumers will tie into the IntelliGrid to allow price signals, decisions, communications, and network intelligence to efficiently flow back and forth. The resulting retail marketplace will offer consumers a wide range of services, including premium power options, real-time power quality monitoring, and home automation services.

Communication Architecture

To realize the vision of the IntelliGrid, standardized communications architecture must be overlaid on today's power delivery system. An integrated energy communications system architecture is needed: a standards-based systems architecture for a data communications and distributed computing infrastructure. Several technical elements will constitute this infrastructure, including data networking, communications over a wide variety of physical media, and embedded computing technologies. The IntelliGrid will enable the monitoring and control of power delivery systems in real time; support deployment of technologies that increase the control and capacity of power delivery systems; enhance the performance of end-use devices that consumers employ; and enable connectivity, thereby revolutionizing the value of consumer devices.

Distributed Energy Resources

Distributed power generation and storage devices, integrated with the power delivery system, offer potential solutions to several challenges that the electric power industry currently faces. These challenges include the need to improve the power delivery infrastructure; provide high-quality power; enable a range of services to consumers; and provide consumers lower-cost, higher-quality power. A key challenge for DER will be to develop ways of integrating these devices seamlessly into the power delivery system, and then monitor and dispatch them that so they can contribute to overall reliability and power quality.

Power Electronics-Based Controllers

In addition to FACTS, other power electronics controllers, based on solid-state devices, offer control of the power delivery system with the speed of a microprocessor. These controllers allow operators to direct power along specific corridors. Power electronics-based controllers can increase power transfer.

Power Market Tools

Market-based mechanisms are needed that offer incentives to market participants in ways that benefit all stakeholders, facilitate efficient planning for expansion of the power delivery infrastructure, effectively allocate risk, and connect consumers to markets. Service providers need a new methodology for the design of retail service programs for electricity consumers, and consumers need help devising ways they can participate profitably in markets by providing dispatchable or curtailable electric loads. Market participants critically need new ways to manage financial risk.

Technology Innovation in Electricity Use

Technology innovation in electricity use is a cornerstone of economic progress. The growth in the gross domestic product over the past 60 years has been accompanied by improvements in energy intensity and labor productivity. Development and adoption of technologies in a number of areas are needed:

- Industrial electrotechnologies
- Indoor air quality

- Lighting
- Recycling processes

The EnergyPort

Assuming that electricity infrastructures are integrated and realizing the ability to connect electricity consumers more fully with electronic communications will depend on evolving a consumer portal to function as a "front door" to consumers and their intelligent equipment. The EnergyPort would sit between consumers and wide-area access networks. The EnergyPort would enable two-way communications between consumers' equipment and energy service entities.

Value-Added Electricity Services

There is enormous consumer advantage in developing and implementing technologies that will enable traditional and nontraditional energy providers and service providers, as well as their customers, to access a variety of electricity-related, value-added services opportunities, including real connecting to electricity markets.

Create ebusiness-enabled opportunities by providing time-saving onsite, online services, based on the integration of electricity with communication systems. These services can be offered by anyone and will improve the digital economy's competitiveness and performance through better integration of power, communications, and delivery of entertainment and emerging wireless communication technologies. Convergence of these technologies and capabilities will create new value within the traditional electricity industry, and remove delays and complexities in decision-making by transforming the current "getting connected" model into a fast and efficient online model. Consumers would enjoy a great deal of benefit if suppliers were able to offer a menu of services related to the "new" electricity.

Electric utilities will be able to expand the portfolio of their business services to include communication, Internet access, real-time online monitoring, and other associated services. The customers will gain real-time access to energy markets and thus better control energy cost and energy utilization. Integrated communication and control will be available to expedite data exchange.

CONCLUSION

There are profound opportunities for innovation to transform the reliability, quality, security, and value of electricity for the 21st century. Conversely, there is a large and growing gap between the performance capability of today's electricity power systems and the needs and expectations of tomorrow's world.

BIBLIOGRAPHY

1. Primen, *The Cost of Power Disturbances to Industrial & Digital Economy Companies*, June 2001.
2. NERC Transmission Loading Relief Logs, www.nerc.com (accessed March 2007).
3. Silberman, S. The Energy Web, Wired magazine, July 2001, http://www.wired.com/wired/archive/9.07/juice.html (accessed March 2007).
4. *SEMI F42-0600 Test Method for Semiconductor Processing Equipment Voltage Say Immunity* (June 2000) and *SEMI F47-0200 Specification for Semiconductor Processing Equipment Voltage Sag Immunity* (February 2000), SEMI.
5. *National Transmission Grid Study*, U.S. Department of Energy, May 2002, http://tis.eh.doe.gov/ntgs/gridstudy/main_screen.pdf (accessed March 2007).
6. Stephen, T.L, *A New Vision for Transmission Operation and Planning Under an Open Power Market*, presented at CIGRE Conference, Montreal, Canada, October 7–10, 2003.

International Performance Measurement and Verification Protocol (IPMVP)☆

James P. Waltz
Energy Resource Associates, Inc., Livermore, California, U.S.A.

Abstract

It is in some ways a strange curiosity that the "microprofession" of measurement and verification has generally ignored the financial side of the question of demonstrating the performance of an energy retrofit project. That is, a component which is conspicuously absent from the International Performance Measurement and Verification Protocol (IPMVP) is any treatment of the conversion of units of energy into dollars saved or costs avoided. It is an unfortunate omission, as conspicuous as the "missing man" in a formation flyover at a memorial service.

This article attempts to fill that hole in the IPMVP by addressing at least five different treatments of establishing a realistic dollar value for the units of energy saved in a performance contract and which are potentially being documented in a measurement and verification process.

INTRODUCTION

The International Performance Measurement and Verification Protocol (IPMVP) is notable in its lack of addressing the valuation of energy savings. This situation may very well be the result of the fact that the vast majority of professionals involved in measurement and verification come from an engineering and/or academic point of view and perhaps have a predilection for dealing with engineering units only. This should probably be considered a significant weakness in IPMVP, as it has been our experience that building owners, in fact care very less about saving energy, but care very much about saving money.

An addition to the IPMVP has been proposed to the technical committee but, for the apparent reasons stated above, was not implemented in the 1997 IPMVP's successor, "IPMVP 2000." This article essentially presents that proposed addition "Valuing Energy Savings," and discusses many different ways of establishing the value of a kilowatt hour saved. In doing so, it will address the following concepts and methodologies:

"avoided cost"
average unit cost
weighted average unit cost
calibrated simulation
real time costing
rate application

Readers should realize that just as *all* methods of measurement and verification are imperfect, yet useful, the various methods of valuing energy savings presented herein are similarly flawed, but also useful.

AVOIDED COST

Facility managers can get into trouble if they do not address the mindset of the "audience" they are working for—generally upper management. As a general observation, upper managers tend not to understand, nor do they particularly care about, the intricacies of energy management and the "wonders" of the engineering world of Btus or kilowatt hours. Rather, they more characteristically tend to think and deal in economic units, i.e., dollars. As a result, trouble can occur due to this difference in mind sets.

As a simple but illustrative example, suppose that the facility manager has managed to decrease energy use by 10% from the prior year. Unfortunately, however, the energy supplier has simultaneously raised his rates by 10%—with the result that the *total cost of energy for the facility has remained the same*. Now, suppose further that upper management, based on what they thought were the promises made by the facility manager for his energy conservation program, decreased the budget for energy by 10%, and has now discovered that they are 10% over budget. Upper management cannot help but conclude that the facility manager has failed to produce the promised "savings" (after all, they paid a million last year and the bill this year is still a million—meaning that *there are no*

☆ This entry originally appeared as "Valuing Energy Savings: Finding the 'Missing Man' in the IPMVP 'Formation'" in *Energy Engineering*, Vol. 101, No. 2, 2004. Reprinted with permission from AEE/Fairmont Press.

Keywords: Performance measurement; International standards; IPMVP; Verification; Energy savings.

savings). The potential consequences for the facility manager are fairly obvious and ominous in this scenario and would be avoided by any facility manager with any sense at all.

Then the astute facility manager, will have educated upper management in such a way that they understand that total energy cost is a product of two components: consumption and *rates*, and that cost may stay the same or even increase at the same time that "savings," better known as "cost avoidance," is being achieved.

This "cost avoidance," then, is the difference between what the actual cost of energy is vs what the cost of energy *would have been* had no conservation actions been taken. In our illustration, while total cost remained the same, it would have increased by 10% had no energy management program been implemented. This concept is particularly poignant in facilities such as acute care hospitals, where a phenomenon known as "energy creep" is fairly common. In many acute care hospitals (and other types of complex facilities as well, of course), over time, more and more diagnostic and other types of electric and electronic equipment is brought into the facility. When this is combined with the natural "wearing down" of the facility's infrastructure systems (lighting, heating, ventilating and cooling (HVAC), etc.), it is not uncommon for a facility's energy use to gradually increase over time. In one specific case documented in the references, a 10-year pattern of growing energy use was observed, approximately 3% for electric and 5% for natural gas, during a period when no active changes to the energy management program were undertaken.

AVERAGE UNIT COST

This is the simplest method of valuing units of energy, and it may be applied to any of the four measurement and verification options (A–D).

This method simply takes the total cost of energy (from the utility bill) and divides it by the total number of units consumed, thereby producing an average unit cost for the units consumed. To set a value for energy units saved, this average unit cost for the current period (typically a month) is simply multiplied by the units of energy the measurement and verification procedures identified as having been saved.

This method is obviously very simple and (perhaps obviously) ignores the fact that energy use often has a time-related value, as embodied, e.g., in time-of-use (TOU) electric rates. Just as some believe that Option A in its simplest form (where few or no actual measurements are made) is a "bogus" measurement and verification (M&V) methodology, the use of average units costs may similarly be disparaged. However, if the energy management program being employed is a broad program that affects energy use in a generally uniform manner over all time periods, then it may well be a perfectly suitable and acceptable methodology to the parties of a performance contract.

WEIGHTED AVERAGE UNIT COST

This is a fairly simple method, which is also applicable to all methods of measurement and verification. It is, however, perhaps best suited to Option A.

This approach makes a number of thoughtful assumptions about the time-occurrence of energy use (or energy *savings*) and performs calculations which apply the appropriate rate schedules to the assumptions to determine a weighted average unit cost for the energy units.

While it is based on assumptions, it may be updated or "calibrated" to actual experience if a time-related energy use pattern data are available (say, in a 15-min demand profile).

Tables 1–3 show three examples of a spreadsheet which performs the calculation of weighted average unit cost. Note that the examples show three different patterns of energy use, and produce distinctly different results based on those patterns. Keeping in mind that the analysis assumes that the load being evaluated is a nominal 1 kW load, then the calculation of annual kilowatt hour, e.g., also equals the equivalent full load operating hours (EFLH—please see the American Society of Heating Refrigeration and Air Conditioning Engineers (ASHRAE) handbook for a discussion of this term if it is one with which the reader is unfamiliar). While a complete explanation of the details of the spreadsheet has not been provided, the majority, if not all, of the calculations can be understood by observation. The author will assist those in need of explanation as required.

REAL TIME COSTING

This method is a bit more rigorous than those discussed above, and is generally oriented towards Option B.

This approach consists of building the rate schedule into the monitoring system and applying the rate schedule "on the fly" as energy savings are (or were) actually occurring.

For example, if a variable frequency drive was applied to a motor which previously operated fully loaded all the time, then the instrumentation, data gathering, and data reduction measurement and verification "package" applied to this retrofit would, during each operating time window (perhaps an hour), measure the then-current

Table 1 Average electrical cost analysis (based on time-of-use (TOU)/savings)

Time of use (TOU) group: 1																**Date: 10/31/02**		
Description: 24 h a day operation																**INIT: HK**		
Period (TOU)		**Time**		**Available hours per day**	**Occ. hours per day**		**Percentage in use (%)**		**Hours of use per day[a]**	**Days per week**	**Weeks per season**	**Days per season[b]**	**Hours of use**	**Kilowatt hour unit cost ($)**	**Total energy cost ($)**	**Peak demand cost ($)[c]**	**Part demand cost ($)[c]**	**Max demand cost ($)[c]**
Season: *summer*																		
Off-peak	M/F	24:00	8:30	8.5	8.5	×	100	=	8.5	5	26.2	128	1088	0.05059	55.04	0	0.00	2.55
Part-peak	M/F	8:30	12:00	3.5	3.5	×	100	=	3.5	5	26.2	—	448	0.05810	26.03	0	3.70	0.00
Peak	M/F	12:00	18:00	6.0	6.0	×	100	=	6.0	5	26.2	—	768	0.08773	67.38	13.35	0.00	0.00
Part-peak	M/F	18:00	21:30	3.5	3.5	×	100	=	3.5	5	26.2	—	448	0.05810	26.03	0	0.00	0.00
Off-peak	M/F	21:30	24:00	2.5	2.5	×	100	=	2.5	5	26.2	—	320	0.05059	16.19	0	0.00	0.00
Off-peak	S/S	00:00	24:00	24.0	24.0	×	100	=	24.0	2	26.0	52	1248	0.05059	63.14	0	0.00	0.00
Off-peak	H	00:00	24:00	24.0	24.0	×	100	=	24.0	NA	NA	3	72	0.05059	3.64	0	0.00	0.00
Summer season totals												**183**	**4392**		**257.44**	**80.10**	**22.20**	**15.30**
Season: *winter*																		
Off-peak	M/F	24:00	8:30	8.5	8.5	×	100	=	8.5	5	26.0	125	1062.5	0.05038	53.53	0	0.00	2.55
Part-peak	M/F	8:30	12:00	3.5	3.5	×	100	=	3.5	5	26.0	—	437.5	0.06392	27.97	0	3.65	0.00
Part-peak	M/F	12:00	18:00	6.0	6.0	×	100	=	6.0	5	26.0	—	760.0	0.06392	47.94	0	0.00	0.00
Part-peak	M/F	18:00	21:30	3.5	3.5	×	100	=	3.5	5	26.0	—	437.5	0.06392	27.97	0	0.00	0.00
Off-peak	M/F	21:30	24:00	2.5	2.5	×	100	=	2.5	5	26.0	—	312.5	0.05038	15.74	0	0.00	0.00
Off-peak	S/S	00:00	24:00	24.0	24.0	×	100	=	24.0	2	26.0	52	1248.0	0.05038	62.87	0	0.00	0.00
Off-peak	H	00:00	24:00	24.0	24.0	×	100	=	24.0	NA	NA	5	120.0	0.05038	6.05	0	0.00	0.00
Winter season totals												**182**	**4368**		**242.06**	**0.00**	**21.90**	**15.30**
Annual totals												365	8760		499.51	80.10	44.10	30.60

(Continued)

Table 1 Average electrical cost analysis (based on time-of-use (TOU)/savings) *(Continued)*

Actual rate schedule		
Rate: PG&E E19S	**Effective: 1/1/98**	
	Summer ($)	**Winter ($)**
Demand charges (*$/kW*)		
Max peak	13.35	0.00
Max part-peak	3.70	3.65
Max demand	2.55	2.55
Energy charges (*$/kWh*)		
Peak	0.08773	0.00
Partial-peak	0.05810	0.06392
Off-peak	0.05059	0.05038
Results of analysis		
Total demand (kW) cost	$155	
Total energy (kWh) cost	$500	
Total cost	$654	
Total cost/kWh	$0.075	
Average cost/kWh w/o demand	$0.057	

All calculations assume a 1 kW load. All demand costs are "percentage in use" times rate in effect for period. Cost calculated herein may be used for valuing consumption or savings.
[a]Hours of use per day = occ. hours per day × percentage in use.
[b]Weekdays "days per season" are less holidays.
[c]Demand cost season sub-totals are for 6 months.
Source: Copyright 1995 Energy Resource Associates Inc. all rights reserved.

Table 2 Average electrical cost analysis (based on time-of-use (TOU)/savings)

Time of use (TOU) group: 3	**Date: 10/31/02**
Description: night lighting from 8 P.M. to 1 A.M.	**INIT: HK**

Period (TOU)	**Time**			**Available hours per day**	**Occ. hours per day**		**Percentage in use (%)**		**Hours of use per day**[a]	**Days per week**	**Weeks per season**	**Days per season**[b]	**Hours of use**	**Kilowatt hour unit cost ($)**	**Total energy cost ($)**	**Peak demand cost ($)**[c]	**Part demand cost ($)**[c]	**Max demand cost ($)**[c]
Season: *summer*																		
Off-peak	M/F	24:00	8:30	8.5	1.0	×	100	=	1	5	26.2	128	128	0.05059	6.48	0	0.00	0.00
Part-peak	M/F	8:30	12:00	3.5	0.0	×	0	=	0	5	26.2	—	0	0.5810	0.0	0	0.00	0.00
Peak	M/F	12:00	18:00	6.0	0.0	×	0	=	0	5	26.2	—	0	0.08773	0.0	0	0.00	0.00
Part-peak	M/F	18:00	21:30	3.5	1.5	×	100	=	1.5	5	26.2	—	192	0.05810	11.16	0	3.70	0.00
Off-peak	M/F	21:30	24:00	2.5	2.5	×	100	=	2.5	5	26.2	—	320	0.05059	16.19	0	0.00	0.00
Off-peak	S/S	00:00	24:00	24.0	0.0	×	0	=	0	2	26.0	52	0	0.05059	0.00	0	0.00	0.00
Off-peak	H	00:00	24:00	24.0	0.0	×	0	=	0	NA	NA	3	0	0.05059	0.00	0	0.00	0.00
Summer season totals												**183**	**640**		**33.82**	**0**	**22.20**	**0.00**
Season: *winter*																		
Off-peak	M/F	24:00	8:30	8.5	1.0	×	100	=	1	5	26.0	125	125	0.05038	6.30	0	0.00	0.00
Part-peak	M/F	8:30	12:00	3.5	0.0	×	0	=	0	5	26.0	—	0	0.06392	0.00	0	0.00	0.00
Part-peak	M/F	12:00	18:00	6.0	0.0	×	0	=	0	5	26.0	—	0	0.06392	0.00	0	0.00	0.00
Part-peak	M/F	18:00	21:30	3.5	1.5	×	100	=	1.5	5	26.0	—	187.5	0.06392	11.99	0	3.65	0.00
Off-peak	M/F	21:30	24:00	2.5	2.5	×	100	=	2.5	5	26.0	—	312.5	0.05038	15.74	0	0.00	0.00
Off-peak	S/S	00:00	24:00	24.0	0.0	×	0	=	0	2	26.0	52	0	0.05038	0.00	0	0.00	0.00
Off-peak	H	00:00	24:00	24.0	0.0	×	0	=	0	NA	NA	5	0	0.05038	0.00	0	0.00	0.00
Winter season totals												**182**	**625**		**34.03**	**0**	**21.90**	**0.00**
Annual totals												365	1265		67.85	0	44.10	0.00

(Continued)

Table 2 Average electrical cost analysis (based on time-of-use (TOU)/savings) *(Continued)*

Actual rate schedule		
Rate: PG&E E19S	**Effective: 1/1/98**	
	Summer ($)	**Winter ($)**
Demand charges (*$/kW*)		
Max peak	13.35	0.00
Max part-peak	3.70	3.65
Max demand	2.55	2.55
Energy charges (*$/kWh*)		
Peak	0.08773	0.00
Partial-peak	0.05810	0.06392
Off-peak	0.05059	0.05038
Results of analysis		
Total demand (kW) cost	$44	
Total energy (kWh) cost	$68	
Total cost	$112	
Total cost/kWh	$0.088	
Average cost/kWh w/o demand	$0.054	

All calculations assume a 1 kW load. All demand costs are "percentage in use" times rate in effect for period. Cost calculated herein may be used for valuing consumption or savings.
[a]Hours of use per day = occ. hours per day × percentage in use.
[b]Weekdays "days per season" are less holidays.
[c]Demand cost season sub-totals are for 6 months.
Source: Copyright 1995 Energy Resource Associates Inc. all rights reserved.

Table 3 Average electrical cost analysis (based on time-of-use (TOU)/savings)

Time of use (TOU) group: 2																		Date: 10/31/02
Description: typical office usage from 7 A.M. to 7 P.M.																		**INIT: HK**
Period (TOU)		**Time**		**Available hours per day**	**Occ. hours per day**		**Percentage in use (%)**		**Hours of use per day**[a]	**Days per week**	**Weeks per season**	**Days per season**[b]	**Hours of use**	**Kilowatt hour unit cost ($)**	**Total energy cost ($)**	**Peak demand cost ($)**[c]	**Part demand cost ($)**[c]	**Max demand cost ($)**[c]
Season: *summer*																		
Off-peak	M/F	24:00	8:30	8.5	1.5	×	100	=	1.5	5	26.2	128	192	0.05059	9.71	0	0.00	2.55
Part-peak	M/F	8:30	12:00	3.5	3.5	×	100	=	3.5	5	26.2	—	448	0.05810	26.03	0	3.70	0.00
Peak	M/F	12:00	18:00	6.0	6.0	×	100	=	6	5	26.2	—	768	0.08773	67.38	13.35	0.00	0.00
Part-peak	M/F	18:00	21:30	3.5	1.0	×	100	=	1	5	26.2	—	128	0.05810	7.44	0	0.00	0.00
Off-peak	M/F	21:30	24:00	2.5	0.0	×	100	=	0	5	26.2	—	0	0.05059	0.00	0	0.00	0.00
Off-peak	S/S	00:00	24:00	24.0	0.0	×	100	=	0	2	26.0	52	0	0.05059	0.00	0	0.00	0.00
Off-peak	H	00:00	24:00	24.0	0.0	×	100	=	0	NA	NA	3	0	0.05059	0.00	0	0.00	0.00
Summer season totals												**183**	**1536**		**110.56**	**80.10**	**22.20**	**15.30**
Season: *winter*																		
Off-peak	M/F	24:00	8:30	8.5	1.5	×	100	=	1.5	5	26.0	125	187.5	0.05038	9.45	0	0.00	2.55
Part-peak	M/F	8:30	12:00	3.5	3.5	×	100	=	3.5	5	26.0	—	437.5	0.06392	27.97	0	3.65	0.00
Part-peak	M/F	12:00	18:00	6.0	6.0	×	100	=	6	5	26.0	—	750	0.06392	47.94	0	0.00	0.00
Part-peak	M/F	18:00	21:30	3.5	1.0	×	100	=	1	5	26.0	—	125	0.06392	7.99	0	0.00	0.00
Off-peak	M/F	21:30	24:00	2.5	0.0	×	100	=	0	5	26.0	—	0	0.05038	0.00	0	0.00	0.00
Off-peak	S/S	00:00	24:00	24.0	0.0	×	100	=	0	2	26.0	52	0	0.05038	0.00	0	0.00	0.00
Off-peak	H	00:00	24:00	24.0	0.0	×	100	=	0	NA	NA	5	0	0.05038	0.00	0	0.00	0.00
Winter season totals												**182**	**1500**		**93.34**	**0.00**	**21.90**	**15.30**
Annual totals												365	3036		203.90	80.10	44.10	30.60

(Continued)

Table 3 Average electrical cost analysis (based on time-of-use (TOU)/savings) *(Continued)*

Actual rate schedule		
Rate: PG&E E19S	**Effective: 1/1/98**	
	Summer ($)	**Winter ($)**
Demand charges (*$/kW*)		
Max peak	13.35	0.00
Max part-peak	3.70	3.65
Max demand	2.55	2.55
Energy charges (*$/kWh*)		
Peak	0.08773	0.00
Partial-peak	0.05810	0.06392
Off-peak	0.05059	0.05038
Results of analysis		
Total demand (kW) cost	$155	
Total energy (kWh) cost	$204	
Total cost	$359	
Total cost/kWh	$0.118	
Average cost/kWh w/o demand	$0.067	

All calculations assume a 1 kW load. All demand costs are "percentage in use" times rate in effect for period. Cost calculated herein may be used for valuing consumption or savings.

[a]Hours of use per day = occ. hours per day × percentage in use.

[b]Weekdays "days per season" are less holidays.

[c]Demand cost season sub-totals are for 6 months.

Source: Copyright 1995 Energy Resource Associates Inc. all rights reserved.

Table 4 Variable air volume cost avoidance

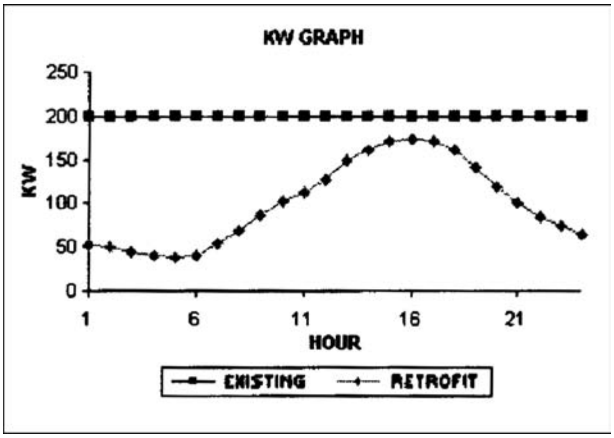
Day type	Weekday cost		Rate sch		Date: 06/09/98	
Hour	Existing		Retrofit		Delta	
	kW/kWh	$/kWh	kW/kWh	$/kWh	kWh	$/kWh
1	200	10.12	52	2.62	148	7.50
2	200	10.12	50	2.51	150	7.61
3	200	10.12	44	2.25	156	7.87
4	200	10.12	41	2.09	159	8.03
5	200	10.12	38	1.94	162	8.18
6	200	10.12	40	2.05	160	8.07
7	200	10.12	54	2.72	146	7.40
6	200	11.62	69	3.99	131	7.63
9	200	11.62	86	5.09	112	6.53
10	200	11.62	103	5.96	97	5.66
11	200	11.62	113	6.55	87	5.07
12	200	11.62	127	7.39	73	4.23
13	200	17.55	148	13.01	52	4.53
14	200	17.55	161	14.11	39	3.44
15	200	17.55	170	14.95	30	2.60
16	200	17.55	173	15.20	27	2.34
17	200	17.55	170	14.95	30	2.60
18	200	17.55	161	14.16	39	3.39
19	200	11.62	141	8.19	59	3.43
20	200	11.62	119	6.89	81	4.73
21	200	11.62	100	5.82	100	6.60
22	200	10.12	85	4.28	115	5.84
23	200	10.12	74	3.73	126	6.39
24	200	10.12	65	3.27	135	6.84
Total/avg.	4800	299.42	2386	163.71	2414	135.71

Rate schedule: PG&E E19S	Effective date: 01/01/98	
	Summer ($)	Winter ($)
Demand charges (*$/kW*)		
Max peak	513.35	0.00
Max part peak	170.0	3.65
Max demand	2.55	2.65
Energy charges (*$/kW*)		
Peak:	0.08773	0.00
Partial-peak	0.05810	0.06392
Off-peak	0.05059	0.05038

(Continued)

Table 4 Variable air volume cost avoidance *(Continued)*

Rate schedule: PG&E E19S	Effective date: 01/01/98	
	Summer ($)	Winter ($)



energy use of the motor, when compared with the original or baseline energy use, calculate the energy units saved, and apply the rate schedule applicable during that window of time.

If a real-time system of data gathering and analysis is employed, then this valuing of energy units saved would also be in real time. However, the approach is just as effective in valuing the energy units saved even if it is applied after the fact to energy data gathered in real time, but analyzed at a later time.

Table 4 shows a sample spreadsheet for an Option B measurement and verification of a variable volume fan conversion. In this example, the baseline conditions have been established by short-term monitoring of the fan motor and a load profile developed as well through short-term monitoring and extrapolation to a year-long load profile by means of linear regression of the short-term load monitoring data and applying the linear regression equation ($Y = mX + b$).

RATE APPLICATION

This method has Option C in mind and assumes that either the utility meter is being employed to measure and record both baseline and postretrofit energy use, or that a meter of similar character has been applied to the facility, and that the baseline energy use has been recorded in such a way that the present rate schedule can be applied to it. That is, the peak demand and energy use in each TOU period is available.

In this instance, the current rate schedule is applied to the recorded energy use and the total cost calculated for both the baseline and postretrofit situations. The cost avoided is simply the difference between the two.

A wealth of proprietary software is already commercially available to implement this methodology, such as Faser, Metrix, and Utility Manager, to name three.

CALIBRATED SIMULATION

This method is not for the faint of heart, and is oriented towards Option D. This method, perhaps due to the extremely wide variation in simulation skill and acumen among its practitioners, has a reputation for being completely hypothetical, extremely effective, or anywhere in between.

Simply put, calibrated simulation requires significant rigor and clear demonstration of its faithful emulation of reality if it is to possess veracity. To achieve this goal, this approach would encompass, e.g., the creation of a computer simulation model of a building or sub-system (say a chiller plant) in its baseline state, and the calibration (and demonstration thereof) of the model to reality. In addition, the retrofitted facility would also be modeled and calibrated as well. In the case of a chiller plant, the reality to which the models are calibrated might be a dedicated utility-style electric meter through which all power consumed by the plant is measured, and other specific measurements, perhaps instantaneous pump and cooling tower electrical demand, total cooling provided by the plant (ton hours output), etc.

Because both the baseline and retrofit conditions are simulated and calibrated, there can be considerable faith, then, in the veracity of the units of energy and the time-related patterns of energy use in the model.

Assuming that the simulation tool used for the modeling has the ability to incorporate and apply the rate schedules in use, both models may be run with whatever current rate schedule is in effect, and very accurately calculate the operation cost of the baseline and the retrofit facilities. Avoided cost, then, is simply the difference between the two.

CONCLUSION

The bottom line of all these efforts, again, is to translate the "technical" determination of units of energy saved into a financial result which management may make use of. Without doing so, energy engineers run the risk of being dismissed as being irrelevant by upper management, so this step in the M&V process is, perhaps amazingly, the one that may be the most important of all!

Investment Analysis Techniques

James W. Dean
Office of Strategic Analysis and Governmental Affairs, Florida Public Service Commission, Tallahassee, Florida, U.S.A.

Abstract
This entry explores various investment evaluation techniques with special attention to their application in energy investments. Key components of the techniques are described, including the concept of the time-value of money, discount rates, and the role of risk and its consideration in performing a financial assessment of a project. Financial evaluations from simple payback periods to internal rate of return (IRR) analysis are discussed with the entry concluding that net present-value analysis is, in almost all cases, the preferred evaluation technique.

INTRODUCTION

You have just been informed that your six-digit lottery ticket hit the $50 million *Big Payout Mega Lottery*. You now are entitled to $5 million per year for ten years or you can take a one-time payment of $30 million. What is the best decision? For the lucky winner, this decision hinges on what economists call the time preference of consumption. An individual could take all the money in a single payout and have the opportunity to immediately acquire goods and services or the individual could defer the opportunity to spend the money and receive larger, future annual payouts. With little knowledge of finance, the newly created millionaire could use a more formal analysis and identify which of the two payout options afford the largest financial return.

While few of us will face such a joyful exercise in our personal finances, everyone who makes financial decisions in their personal or business environment will either implicitly or explicitly have to make a decision like our lottery winner. Individuals who consider the rated miles per gallon expectations in purchasing a car or who consider the energy-efficiency rankings in purchasing a new appliance are in some fashion doing a financial analysis. Likewise, a corporation or business has a fiduciary responsibility to make investments that maximize the overall return to the firm. Many financial decisions involve multiple time periods and uneven cash flows, and without some analytical tools to simplify and compare options, individuals and companies would be stymied in their efforts to estimate the total net worth of investments or to select from competing opportunities that would maximize their overall monetary gain.

In the first part of this entry, two fundamental concepts used in performing financial analyses will be introduced: the time-value of money and the discount rate. The latter portion of the entry explores a variety of financial techniques that are commonly used in evaluating and ranking the value of investment opportunities when cash flows occur over time. These kinds of applications are commonplace in energy investments. Examples include investments in oil and gas drilling projects, power plants, transmission lines, natural gas pipelines, and energy-efficiency investments. In the energy industry in particular, many projects require substantial upfront capital and the associated revenue stream generally occurs over long time periods. However, the investment evaluation techniques discussed here are generic and are routinely used in most applications. Finally, the two most commonly used analytical techniques—present-value analysis and IRRs—are discussed and their relative advantages and disadvantages are examined.

FUNDAMENTALS OF INVESTMENT EVALUATION

It is critical to appreciate that the value of money is time dependent. Money received now has more value than money received in the future as long as there is a positive interest rate that an investor can earn. The proof of this statement is intuitive. If given some amount less than a dollar today, one could, with a positive interest rate, have a dollar in the future. Thus, a dollar today cannot be equal to a dollar in the future given a positive interest rate environment. Developing a comparable value for time disparate dollars is done using a variety of present-value techniques.

One of the most commonly employed investment metrics is the simple payback period. Its strength is in its

Keywords: Present-value analysis; Internal rate of return; Discount rate; Simple payback; Variable discount rates; Net present value analysis.

computational simplicity. The payback period is the number of years (or months) until the revenue stream equals or exceeds the initial project outlay. This technique is frequently used in advertisement to convince potential customers how shrewd a purchase might be. For example, assume a consumer were to spend $500 for the installation of attic insulation to make his house more energy efficient. If this investment saves $100 a year in energy costs, then the simple payback period would be five years. Payback analysis does permit a relative ranking of project worthiness when multiple projects are being evaluated. In other words, projects with shorter paybacks would tend to be superior to longer-dated projects.

However, the simplicity of the technique is offset by two important omissions. The simple payback does not account for the time-value of money nor does it assign any value to the positive cash generated after the breakeven period is reached. In addition, if projects have cash flows that switch from negative to positive and back, even relative rankings are obscured when comparing projects.

Payback analyses can be improved upon by calculating a current value for the future cash flows up to the payback period. Doing this typically results in a longer payback than the simple payback. This approach more accurately reflects the value of the capital investment that has been "tied up" in the project. However, properly valuing the time dependency of the cash flows against the initial investment outlay does not remedy the other problem of correctly accounting for cash flows beyond the breakeven point.

The correct solution to this problem entails using present-value analysis for the life of the entire project. Present-value analysis offers a powerful analytical tool to compare cash flow streams of both inflows (revenue) and outflows (expenditures) and standardizes the magnitude and timing of the flows such that the values are represented in today's currency values. The term "current dollars" refers to values that have been standardized to reflect values in today's dollars and "nominal dollars" refers to the value in the future periods under consideration. When present-value analyses incorporate the value of any initial capital expenditure, and year-by-year inflows and outflows are netted out, the final valuation constitutes a "net present value" (NPV) of the overall project's worth.

In the simple case, a one-year present-value term can be represented by:

$$PV = CF_1/(1 + r_1)$$

where CF_1 is the cash flow in time period 1 and r represents the rate of return that the investor is willing to accept for an alternative investment with comparable risks.

By extension, the present value of multiple years of cash flows can be represented by:

$$PV = \frac{CF_1}{(1 + r_1)^1} + \frac{CF_2}{(1 + r_2)^2} + \frac{CF_3}{(1 + r_3)^3} + \cdots + \frac{CF_t}{(1 + r_t)^t}$$

This equation can be collapsed into:

$$PV = \sum_{t=1}^{N} \frac{CF_t}{(1 + r)^t}$$

This basic formula can be extended to include capital outlays. With the addition of any initial capital outlay CF_0, this present-value analysis becomes a NPV of the form:

$$NPV = CF_0 + \sum_{t=1}^{N} \frac{CF_t}{(1 + t)^t}$$

The CF_0 term is only the initial outlay divided by $(1+r)^0$, which equals the outlay divided by one. However, multiyear outlays are likely to occur, especially on large construction projects. In this case, the outlays are likewise discounted and summed along with future cash inflows. Thus, the NPV formula can be further reduced to account for the initial and subsequent time periods, resulting in:

$$NPV = \sum_{t=0}^{N} \frac{CF_t}{(1 + t)^t}$$

Each CF_t represents the balance of cash flow for each time period. Positive flows are offset by any outflows such as maintenance, labor, accrued interest, or any other cost. Negative flows should be offset by projected revenues, impact of taxes, depreciation, and any terminal salvage value.

For most capital budgeting analyses, an annual end-of-year convention is used to simplify the timing of receipts or expenditures. However, projects involving large capital expenditures or large cash inflows frequently incorporate more discrete time periods for booking cash flows. This is especially true for multiyear, capital-intensive projects that incur substantial interest and construction costs prior to generating revenue. Examples in the utility industry include power plant and refinery construction schedules. Because hundreds of millions of dollars can be spent over the course of a year, accrued interest during construction can be substantial. Thus, when comparing multiple projects that have large intrayear cash flow differences, more fractional time periods provide greater accuracy of total project costs.

If one uses an end-of-year accounting convention, then the cost of funds during construction is assumed to be zero and interest begins to accrue at the beginning of the subsequent year. Conversely, if all cash flows are assumed to occur at the beginning of the year, then accrued interest is overstated. The analyst may elect to do mid-year expenditures to average these out. For projects that require greater scrutiny or for managing actual construction budgets, more discrete time periods such as months or quarters can be used as the unit of analysis. Of course, the discount rate should be adjusted to reflect the compounding impacts of the more discrete time periods.

Another complexity involved with cash inflows and outflows is ensuring that inflation is appropriately accounted for in the analysis. In performing net present-value analysis, there are two strategies that can be used to adjust the cash flow streams. As previously mentioned, dollars are accounted for in either nominal or real terms. Nominal dollars measure cash flow including the impact of inflation whereas real dollars adjust for inflationary expectations. One can convert the nominal inflows and outflows into real cash flows by dividing by an anticipated inflation factor. In this case, a real rate of return would be used as the discount factor in the NPV formula. Alternatively, the cash flows could be stated in nominal dollars, but the value of r would be adjusted higher to account for inflation-related risk.

Finally, it is important to note that any initial cash outlay represents incremental project costs. For example, if special equipment has been purchased or if buildings have been constructed to undertake a project but the project is later abandoned, then such costs are called “sunk costs” in economic jargon. If a subsequent project is undertaken, the initial capital outlays should not include any unrecoverable, sunk costs. However, any additional first-year outlays and all future cash flows should be included in the analysis. If these previously incurred costs have residual value and could be sold or leased to generate cash inflows, then these values would be reflected in the initial year’s outlay. If these costs have no recoverable value, regardless of what the initial costs were, then for purposes of evaluating project investments, their outlay value is zero when used in subsequent project evaluations.

DETERMINING THE APPROPRIATE DISCOUNT RATE

In business applications, the r term is referred to as the discount rate or cost of capital and is a rather complex concept. At its most basic, the discount rate embodies all known sources of risk and therefore reflects what investors must anticipate earning to assume those risks. Thus, as project risk increases, the discount rate used in evaluating any specific project must also go higher.

Examples of project risk abound. In inflationary investment periods, the investor must be compensated for the loss of purchasing power between time periods. Second, the discount rate must capture the relative risks associated with a particular investment decision. Examples include uncertainty of cash inflows, uncertainty related to completing the project, and uncertainty surrounding the political and legal climate in which the project will operate. Opposition to the construction of large infrastructure projects such as liquefied natural gas terminals or power plants is common. Prudent investors will require a greater return to lend capital to finance such projects if the possibility exists that the project will not be completed and thus incur uncompensated development costs, or on the other extreme, if the project is completed but a change in political regimes leads to the nationalization of the project by a foreign government. These types of risks are typically referred to as project-specific risks.

Moreover, the appropriate discount rate is almost unique in that it will vary depending on if the investor is an individual, a business, or a government agency. Within these subgroups, individual circumstances will further define the preferences to defer consumption and the willingness to assume other risks. For many individuals, the need or preference for immediate consumption results in very high discount rates. In the case of individual consumers investing in conservation and energy-efficiency technologies, discount rates vary inversely with income and for individuals with limited capital personal discount rates can be extraordinarily high.

Businesses include a number of components in assessing project risk. Projects are examined for intrinsic or project-specific risks, as discussed above. In addition, the project will be examined in the context of all the firm’s projects and in the context of other market opportunities. Thus, determining what is the appropriate discount rate to use is conceptually difficult, even if the arithmetic computation of the net present-value formula is straightforward.

One factor that goes into the development of the appropriate discount rate is a firm’s cost of capital. For businesses that use a combination of equity and debt to finance projects, the weighted average cost of these component sources of capital typically reflect the individual company’s discount rate. This assumes, however, that any individual business has a risk profile identical to that of the industry as a whole. In other words, if your “riskiness” is exactly the mean riskiness of comparable businesses in your industry sector, then a starting point for assessing individual projects would be the firm’s cost of capital. Any specific project would need additional scrutiny to account for any unique project risks and the appropriate discount rate that should be used in the analysis could be higher or lower than the firm’s cost of capital.

Finally, because debt is frequently a tax-deductible expense, the debt component should be adjusted for tax savings. The following formula would result in an approximate after-tax cost of capital:

$$\text{Weighted after tax cost of capital} = \frac{r_D(1 - T)D + r_E E}{D + E}$$

where r_D is the required interest on debt, T is the appropriate marginal tax rate, D is the amount of debt that is financed, r_E is the expected return on equity, and E is the equity value. It is important to note that the marginal cost or last-source cost of capital is the correct factor to use as the discount rate for new investments. This discount rate would be the appropriate rate to use in any present-value analysis as long as the various projects under consideration possessed similar risk characteristics. As a convention throughout this entry, discount rate and cost of capital are used synonymously.

ALTERNATIVE INVESTMENT METRICS

Net present-value analysis provides a final value on a project-specific basis. Generally, a positive NPV should be undertaken and a negative project should not. However, given that capital is limited and not all investment opportunities afford the same return, alternative investment criteria can be useful in comparing projects. The discussion below should lead to caution when these alternative techniques are used without consideration of the net present-value results. Decision makers can end up with confused if not outright misleading results if these alternative techniques are used in isolation.

Many finance professionals, especially those in the public sector, rely on benefit-to-cost analyses as analytical tools. Most investment opportunities involve both cash outflows and inflows. When present-value analysis is applied and the annual outflows and inflows are not netted out but are instead simply discounted back as current dollars, then the ratio of the present-value benefits (inflows) divided by the present value of costs (outflows) gives a relative ranking against which multiple projects can be judged. In corporate finance, such rankings are referred to as profitability indexes.

Alternatively, the IRR is another commonly used investment metric. Most corporate finance texts define the IRR as that discount rate which equates the present value of a project's expected cash inflows to the present value of the project's cash outflows. The resultant percentage, when applied to the NPV equation, would result in a zero present value for the project.

In terms of the actual formula for IRR, the fundamental net present-value equation is set to zero and IRR notation replaces r for the discount rate.

$$0 = \sum_{t=0}^{N} \frac{CF_t}{(1 + IRR)^t}$$

Thus, IRR answers the question of what is the internally generated return on investment when assuming a project's cash flows result in a breakeven project. In one sense, IRR is a measure of profitability based only on the timing and magnitude of internally generated cash flows of the project. It is not the same as the discount rate the company has determined is needed to initiate the project. Only when a project results in a zero NPV will those two terms be equal.

Frequently, corporate finance professionals prefer the IRR metric because the concept is familiar to shareholders, directors, senior managers, and external investors. However, the use of internal rates of return as investment criterion warrants some caution. Prior to the proliferation of handheld business calculators and spreadsheets, the value of IRR was largely calculated by iterative estimation techniques. Depending on the scope and duration of the cash flow streams, this exercise could be rather burdensome. Spreadsheets and handheld calculators have eliminated these practical difficulties. Interpreting and acting upon IRR results is still problematic, however. Internal rates of returns when considered in isolation can lead to some erroneous decisions on the profitability ranking of different investment opportunities.

It is important to understand how IRR analyses relate to present-value analyses. Present-value analysis specifies the discount rate or cost of capital a priori. This rate is based on external factors such as prevailing interest rates and the expected return on invested equity for projects with similar risk attributes. Thus, any positive present value would imply a positive return on invested capital and indicates the project should be undertaken. The calculation of an IRR asks a different question. Given the expected positive and negative cash flows resulting in a zero present value, what would be the implied return on invested capital?

There are two basic scenarios to consider. Investment capital is frequently, but not always, limited. With adequate funds, such that any two or more projects can be undertaken, an IRR greater than the cost of capital and a positive present-value analysis lead to the same conclusion—undertake the projects. Most firms and individuals do not have unlimited funds. In these cases, selecting one project eliminates the opportunity to undertake the second one. These are referred to as mutually exclusive projects. With mutually exclusive cases like these, the two metrics can result in, at best, ambiguous decision criteria and, at worst, conflicting investment-decision rules.

A NPV AND IRR COMPARISON

Let us explore how the NPV and IRR results can lead to potentially conflicting investment signals. Assume that two project opportunities are available. They have the following cash flows:

Year	Project A ($)	Project B ($)
0	−5000	−5000
1	2500	500
2	1500	1000
3	1000	1250
4	800	1800
5	500	2500

As the analyst assigned to evaluate the two projects, you are told that the company's cost of capital is 5%. You are directed to prepare an NPV and IRR analysis and to make a recommendation to the board of directors. The handy spreadsheet functions indicate that at 5% cost of capital, the NPVs for Project A and B are $624.06 and $859.71, respectively. The IRR function on the spreadsheet produces an IRR of 11% for Project A and 10% for Project B. Because both projects have positive NPVs and the IRR is greater than the cost of capital, the recommendation should be to undertake both projects. In this example, project feasibility is independent. In other words, capital is available to complete both opportunities.

The ambiguity occurs if these are mutually exclusive projects. Typically, this implies capital is not available to do both. While Project B has the higher NPV, its IRR is lower than Project A's. Which recommendation should be made to the board, and more importantly, how do you explain this anomaly as part of the PowerPoint presentation?

A graphic presentation of the relationship between discount rates and NPVs can help explain the apparent conflict. The table below calculates the NPV for these two projects from a zero discount rate to 15% using increments of 5%.

Cost of capital (%)	Project A NPV ($)	Project B NPV ($)
0	1300.00	2050.00
5	624.06	859.71
10	109.62	1.69
15	−285.53	−621.81

Next, the NPVs are plotted against the cost of capital and presented in Fig. 1 below.

A number of characteristics about the two investment projects can be gleaned from this chart. First, notice that at the lower range for the cost of capital, Project B has higher NPVs—simply because of the higher cash inflows and these inflows losing little time value due to the lower discount rates being applied. The two project NPV streams intersect and cross over when the discount rate is approximately 8.2%. At this rate, the NPVs for both projects are approximately $279. For any discount rate above this level, Project A has a higher NPV and would be the preferred project until it turns negative around 11.3%. Project B turns negative around 10% while Project A still has a positive $109 NPV.

The internal rates of return for Project A and Project B are 11 and 10%, respectively. Recall that the NPV calculation specifies the firm's actual cost of capital for the project. If the NPVs are positive and the IRR is greater

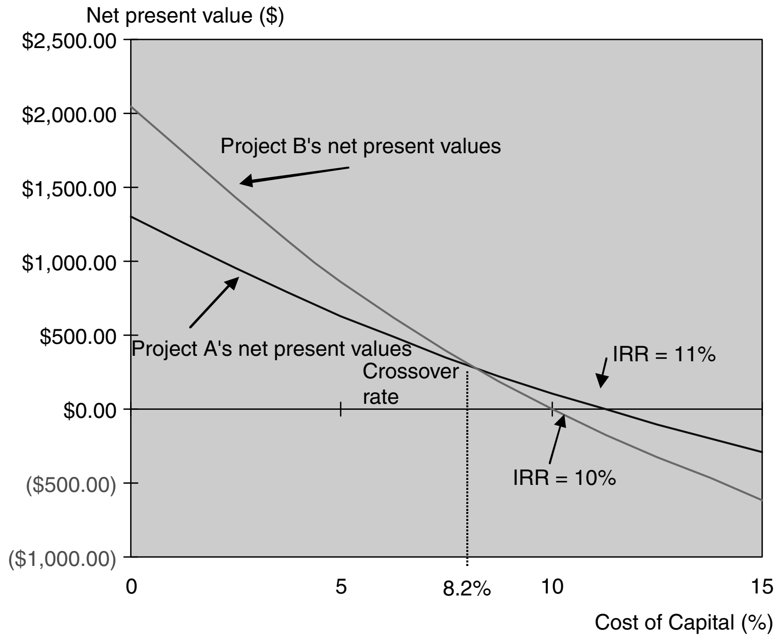


Fig. 1 The net present values of two projects.

Inv–Light

than the underlying cost of capital, then both projects yield positive returns and should be undertaken. If the actual cost of capital is 10%, then as previously noted, Project B is only marginally cost effective. If the actual cost of capital was 10.1%, Project B's NPV turns to a -$12. At this point, the 10% IRR is lower than the cost of capital and Project B should not be undertaken. Again, Project A's IRR of 11% exceeds the required cost of capital and remains a viable investment.

Let's return to the limited capital or mutually exclusive case. If only one of the projects can be undertaken, which one is the superior selection? If both IRRs exceed the required cost of capital, then the decisions should default to the project with the highest NPV. This option maximizes the return to the firm. The ambiguity occurs when the actual cost of capital is less than the crossover or intersection rate. The graph indicates that in all such cases, the NPVs for Project B are higher than for Project A. Yet, the IRR is higher for Project A. If one were explaining the selection before the board of directors, would Project A with the higher IRR not be the more defensible investment?

The answer is no. These mixed investment signals are being generated by the forced computation of the IRR and the assumed reinvestment rate of cash proceeds from the two projects. Because IRR analyses force a discount rate to result in a zero NPV, the formula results in a high rate that is necessary to discount the large, front-loaded cash flows from Project A back to zero to offset the initial capital outlay. Implicit in this process is the assumption that the earlier cash proceeds are reinvested at the IRR. The net present-value method assumes that the proceeds are reinvested at the firm's cost of capital. Therefore, the back-end-loaded cash stream generated by Project B is being discounted at a lower and more realistic cost of capital rate that results in a higher NPV. As mentioned earlier in this exposition, the discounting of the magnitude and timing of the cash flows is causing these divergent results between IRR and NPV analyses.

Table 1 presents two cash flow patterns that dramatically indicate how the IRR can lead to the selection of an inferior project when exclusivity exists.

The NPV of Project A is $1715 when a 10% cost of capital figure is used. At 10%, Project B has an NPV of $1872. With no other information, Project B would be the preferred option. However, the IRRs lead to a different and erroneous conclusion. Project A has an IRR of 23% and Project B yields only a 15% IRR. Here again, the timing and magnitude of cash flows are distorting the two criteria because the mathematical solution of deriving the IRR forces a higher rate to be used to discount the early positive cash flows in Project A back to zero.

VARIABLE DISCOUNT RATES

Another confounding problem when calculating IRR is when the discount rates vary over the project horizon. This can occur when the cost of capital varies because of differences in the yield curve, or term structure, of debt. Because the IRR is derived as a single point value using iterative techniques, the calculated value must be the same for each year of the analysis. However, companies may have a different cost of debt for financing the early years of a project vs. the cost of capital for the later year financing requirements. For example, low cost, short-term commercial paper may be used for the early year's financing on a project and higher cost bonds for the remainder of the project's life. Internal rate of return techniques cannot easily handle this complexity. Other factors that might change the discount rate are phased in changes to the tax rates or depreciation schedules.

Another odd result can occur when projects have cash flows that switch between positive and negative values. Sometimes referred to as nonnormal cash flows, applying IRR techniques can result in more than one IRR being generated. This is likely to occur when a project generates a positive NPV for some portion of the project life and then the cash outflows generate a negative NPV value for several years and then reverse back to positive. The analyst can end up with two return rates and little confidence in which rate accurately captures the true IRR. Net present-value analyses can address issues like these by permitting time varying discount rates.

CONCLUSIONS

It should be evident from the discussion above that net present-value metrics have a variety of features that make them preferable to IRR analysis. They are computationally more reliable while providing a clearer indication of the value that a specific project generates for the bottom line.

Table 1 NPV and IRR comparison-example cash flows

Project	CF_0	CF_1	CF_2	CF_3	CF_4	CF_5	CF_6	CF_7	CF_8	CF_9	CF_{10}
A	−9000	6000	1000	3000	150	150	150	150	150	150	150
B	−9000	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800

The conventional decision rule states that a firm should undertake all projects with positive NPVs and IRRs that are greater than the discount rate. However, in situations where projects are mutually exclusive, IRR should be carefully evaluated before using it as the primary indication of a project's profitability. Despite the widespread use of IRR, practitioners of investment analyses should employ this technique only as a secondary evaluation tool to net present-value strategies.

This entry has provided an overview of the mechanics of net present-value analyses and discussed several of the conceptual issues involved using these techniques. However, such a brief treatise does a disservice to novices using these evaluation techniques. The description and examples all have yielded a single NPV that belies major simplifications of assumptions and ignores the single most difficult factor to quantify—the project's risk. For example, we assumed that the cash flows for any project were known with certainty; we assumed the cost of capital was constant over the project's life; and we assumed any two projects had identical risk profiles. Such certainty allows a calculated, single-point NPV estimate to be confidently taken to lenders to seek financing for the project.

In real applications, such certainty rarely exists. As a consequence, a toolbox of additional analytical techniques can be incorporated into NPV analysis to give investors a more complete assessment of project parameters. Most of these tools are designed to better quantify and thus illuminate the risk associated with investments. For example, recent work in option theory allows option value to be incorporated into the project's pro forma. This may be particularly important when early development work is being undertaken with uncertain cost factors such as volatile interest rates or wide ranging estimates in construction costs. The ability to proceed, terminate, or transfer a project prior to a substantial commitment of capital can increase project flexibility and thus can impart real value to the project's overall worth.

Along these same lines, practitioners frequently will use sensitivity analyses to change the value of critical variables to assess the impact on the NPVs. Such efforts result in a series of NPVs under varying conditions such as high and low interest rate environments. Even more sophisticated Monte Carlo techniques run literally thousands of combinations of factors to develop a statistical distribution of NPVs for any given project. With Monte Carlo distributions, profitability is described in terms of probable payouts instead of single-point estimates. We mention these tools to remind readers that while NPV is the linchpin for investment analysis, it is only the starting point.

BIBLIOGRAPHY

1. Brigham, E.F.; Gapenski, L.C. *Financial Management*, 7th Ed.; Dryden Press: New York, 1994.
2. Hausman, J.A. Individual discount rates and the purchase and utilization of energy using durables. Bell J. Econ. **1979**, *Spring*, 33–54.
3. Myers, S.; Brealey, R. *Principles of Corporate Finance*, 6th Ed.; McGraw-Hill: New York, 2000.
4. Osteryoung, J.; McCarty, D.E. *Analytical techniques for financial management*; Grand Publishing, Inc.: Columbus, U.S.A., 1980.
5. Smith, G. Nominal and real required returns in present value analysis. Eng. Economist **1988**, *33* (4), 331–347.
6. Train, K. Discount rates in consumers energy-related decisions: A review of the literature. In *The Economics Energy*, Vol. 1; Elgar Press: Cheltenham, U.K., 2000; 385–395.

LEED-CI and LEED-CS: Leadership in Energy and Environmental Design for Commercial Interiors and Core and Shell

Nick Stecky
NJS Associates, Denville, New Jersey, U.S.A.

Abstract
The U.S. Green Building Council (USGBC) has followed up on its very successful Leadership in Energy and Environmental Design for New Construction (LEED-NC) rating system with two new systems—one for Commercial Interiors (CI) and the other for Core and Shell (C&S). (U.S. Green Building Council (www.USGBC.ORG) LEED green building rating systems for Commercial Interiors and Core and Shell).[1] These two companion programs, LEED for C&S and CI, were developed in response to marketplace demand for a LEED rating system that would enable tenant organizations to have a green workplace while enabling developers the opportunity to provide green buildings for tenants to move into. In the commercial building developer marketplace as currently exists, developers can build a basic building consisting of a C&S with few to no interior fitouts. As tenants are secured, interior fitouts are constructed in accordance with tenant requirements. The result of this process is that the developer can control the greenness of the C&S, and the tenant can control the greenness of the fitout spaces, but neither of them controls the whole building design. Leadership in Energy and Environmental Design for New Construction was unable to accommodate effectively the need to "unbundle" the various elements of a green building—hence, the development of C&S and CI. This entry concludes with a short discussion of some of the obstacles that the USGBC and the LEED programs have met and overcome, and with a forecast of LEED activities.

SUMMARY INTRODUCTION OF TWO NEW LEED RATING SYSTEMS—LEED-CS AND LEED-CI

LEED-CI for Commercial Interiors and LEED-CS for Core and Shell

Leadership in Energy and Environmental Design for Core and Shell (C&S) and Commercial Interiors (CI) were developed in response to marketplace demand for a LEED rating system that would enable tenant organizations to have a green workplace while enabling developers the opportunity to provide green buildings for tenants to move into. In the developer marketplace as currently exists, developers can build a basic building consisting of a C&S with few to no interior fitouts. As tenants are secured, interior fitouts are constructed in accordance with tenant requirements. The result of this process is that the developer can control the greenness of the C&S, and the tenant can control the greenness of the fitout spaces, but neither of them controls the whole building design. Thus, it becomes difficult to apply Leadership in Energy and Environmental Design for New Construction (LEED-NC)—which is based upon whole-building design, interiors as well as C&S—to the commercial developer's building process. The USGBC recognized this dilemma and developed the LEED-CI and C&S rating systems based upon LEED-NC. Consider CI and C&S to be an unbundling of the NC rating system that assigns responsibilities for LEED credits to the entity that has control over those sections. For example, the developer will be responsible for site location, site work, building envelope, central Heating Ventilation and Air Conditioning (HVAC), energy-consuming systems, etc. The tenant will be responsible for the interior spaces, products, office lighting, furniture, paints, office HVAC, etc.

One may ask: how is it known which system to apply, NC or the combination of CI and C&S? The USGBC response is to use NC when the owner is the occupant and has control over the C&S as well as the interior spaces. Note that the two systems, CI and C&S, were developed for circumstances wherein the owner/developer is a different entity from the tenants or occupants. The USGBC goes further in its discussion of this question to include a clarifying statement that C&S and CI are not an option for owners who wish to achieve exemplary performance in the C&S while failing to meet LEED standards in the buildings' fitout. Applicability will become clearer after reading the actual LEED credits per system later in this entry. Note also that throughout this entry, the term LEED credits will be used interchangeably with LEED points.

Keywords: Leadership in Energy and Environmental Design (LEED); Sustainability; Energy efficiency; Indoor air quality; LEED-CI; LEED C&S; LEED Commercial Interiors; LEED Core and Shell.

LEED-CI

Leadership in Energy and Environmental for Commercial Interiors addresses the specifics of tenant spaces primarily in office, retail, and industrial buildings. It was formally adopted in the fall of 2004. The companion rating, Leadership in Energy and Environmental Design for C&S, is currently being balloted, and adoption is expected in the fall of 2006. Together, LEED-CI and LEED-CS will establish green building criteria for commercial office real estate for use by both developers and tenants.

Leadership in Energy and Environmental Design for Commercial Interiors serves building owners and occupants, as well as the interior designers and architects who design building interiors and the teams of professionals who install them. It addresses performance areas including water efficiency, energy efficiency, HVAC systems and equipment, resource utilization, furnishings, and indoor environmental quality.

Benefits of LEED-CI

A number of potential benefits arise out of LEED-CI, many of which are results of the CI focus on providing ideal indoor environments. Included are elements such as daylighting, temperature control, improved ventilation, and adherence to American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Standard 62.1-2004, Ventilation for Acceptable Indoor Air Quality[2] and Standard 55, Thermal Environmental Conditions for Human Occupancy.[3] Thus, CI provides for occupant well being and productivity. It can also aid in retaining employees and reducing absenteeism. Liability due to poor indoor air quality is reduced. Marketability is increased; churn costs are reduced; and maintenance and operations costs are reduced. The result is an optimum workplace environment that can generate large savings for management.

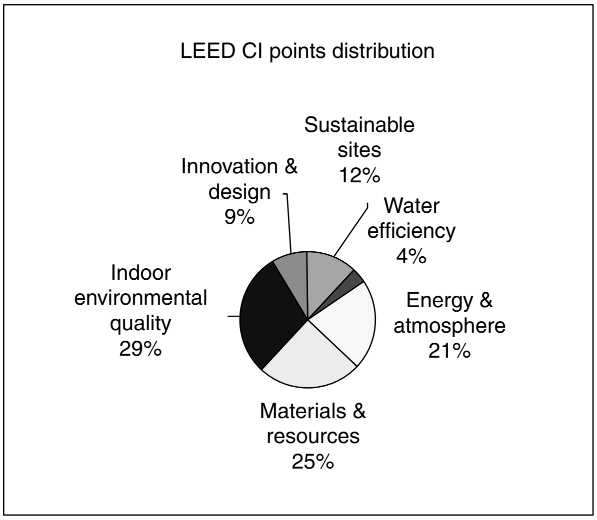
LEED for Commercial Interiors (CI)

- Addresses the design and construction of interiors in existing buildings and tenant fitouts in new C&S buildings
- Pilot: 2002–2004
- Achievements: more than 45 projects in pilot
- LEED-CI adopted in fall 2004.

LEED-CI Point Distribution

Note the same five basic categories as the LEED-NC rating system, but with different weighting from NC. Energy and Atmosphere is the largest point category in NC, but it ranks only third in CI points contributions. This is as expected because CI's influence is primarily on interior spaces, and the interior professionals (such as interior designers and architects) have the major influence.

LEED-CI maximum possible points per category	
Sustainable sites	7
Water efficiency	2
Energy and atmosphere	12
Materials and resources	14
Indoor environmental quality	17
Innovation and design process	4
LEED accredited professional	1
Total points available	57



As in LEED-NC, there are four levels of LEED-CI certification:

Certified	21–26
Silver	27–31
Gold	32–41
Platinum	42–57

LEED-CI—A Technical Review

The LEED formats for rating a green building consist of:

- Prerequisites. These are mandatory requirements, and all must be satisfied before a building can be certified.
- Credits. Each credit is optional, with each contributing to the overall total of credits. This total is used to

determine the level of rating that a building will earn, ranging from Green at the lower level to Silver, Gold, and Platinum.

Sustainable Sites. The maximum number of possible credits is 7. The following credits are very similar to what was described in the LEED-NC section.

Prerequisites. None in this category.

Credits

- Site Selection. Up to 3 credits available; requires the selection to be a LEED Certified Building or to locate the tenant in a building that has in place two or more of the following: Brownfield Development, Stormwater Management Rate and Flow, Stormwater Management Treatment, Heat Island Reduction for roof or nonroof, Light Pollution Reduction, Water Efficient Irrigation, Innovative Wastewater Technologies, Water-Use Reduction, Onsite Renewable Energy, and other quantifiable environmental performances. The obvious choice for the tenant entity that wants to be in a LEED-CI facility is to locate in a LEED-NC or C&S building. Otherwise, the multiple requirements for the alternative path to these credits can become difficult to satisfy without a great deal of participation on the part of the owner.
- Development Density and Community Connectivity. 1 credit available. The intent is to channel development to urban areas with existing infrastructure, protect greenfields, and preserve habitat and natural resources.
- Alternative Transportation. Total of 3 credits available through three categories:
  - Proximity to commuter, light-rail, subway station, or bus transportation
  - Provide for bicycle commuting by requiring secure bicycle storage and convenient shower/locker facilities for 5% or more of the tenants
  - Limit the number of parking spaces for single-occupancy vehicles and provide priority for vans and carpooling

Water Use Efficiency. Only 2 credits available.

Prerequisites. None in this category.

Credits

- Water-Use Reduction. 1 point. Requires tenant to use 20% less water in aggregate than the water-use baseline calculated for the tenant space after meeting the Energy Policy Act of 1992 fixture performance requirements.
- Water-Use Reduction. 1 additional point if the water-use reduction is 30% instead of the 20% listed in the first credit.

Energy and Atmosphere. Up to 12 credits available.

Prerequisites. Three required.

- Fundamental Commissioning. Focus on HVAC and R, lighting, daylighting and associated controls, renewable energy systems, and domestic hot water systems.
- Minimum Energy Performance. Comply with American Society of Heating, Refrigerating and Air Conditioning Engineers/Illuminating Engineers Society of North America (ASHRAE/IESNA) Standard 90.1-2004[4] or the local energy code, whichever is more stringent.
- Chloro Flouro Carbon (CFC) Reduction in Heating Ventilation and Air Conditioning and Refrigeration (HVAC&R) Equipment. To reduce ozone depletion, require zero use of CFC-based refrigerants in tenant HVAC&R systems.

Credits

- Optimize Energy Performance Lighting Power. Up to 3 credits. The first credit requires the lighting power allowance to be 15% less than the ASHRAE/IESNA baseline standard. The second credit is for 25% better than the standard, and the third credit is for 35% or better than the standard.
- Optimize Energy Performance Lighting Controls. 1 credit available; requires use of daylight sensors to control lighting levels within 15 ft of windows or skylights.
- Optimize Energy Performance HVAC. Up to 2 credits. There are two options for compliance. The first relates to equipment efficiency, as described in the New Buildings Institute, Inc. publication "Advanced Buildings: Energy Benchmark for High Performance Buildings." In addition, provide appropriate zone controls for temperature control. The second way to compliance is to demonstrate that HVAC system-component performance criteria for tenant spaces are 15% better than the ASHRAE/IESNA standard for 1 credit. An additional credit is given if the performance is 30% better than the ASHRAE/IESNA standard.
- Optimize Energy Performance Equipment and Appliances. Up to 2 points. This applies to all Energy Star-eligible equipment installed in the project, including appliances but excluding HVAC and lighting and building envelope products (these were covered in sections above), and earns 1 point if 70% of applicable equipment is Energy Star rated. The second point is earned if 90% of applicable equipment is Energy Star rated.
- Enhanced Commissioning. Earns 1 point. In addition to the Fundamental Commissioning prerequisite, this credit requires additional commissioning to do design review before construction, to review the energy-related systems, to develop a manual for recommissioning, and to verify that training requirements for operations personnel are completed.
- Energy Use. Up to 2 points; provides two case paths to compliance. Case A requires submetering for energy

use in the tenant spaces and negotiating a lease wherein the tenant pays the energy bills, which are not included in the rent. The purpose of this is to highlight to the tenant the actual costs of energy and thereby encourage conservation. In the alternative Case B, install continuous metering for energy end uses such as lighting systems and controls, constant and variable motor loads, Variable Frequency Drive (VFD) applications, chiller efficiencies at variable loads, cooling loads, boiler efficiencies, and more.

- Green Power. 1 point; requires that at least 50% of the tenant's electricity come from renewable sources by engaging in at least a 2-year energy contract.

Materials and Resources. Up to 14 possible credits in this category.

Prerequisites. One required in this section.

- Storage and Collection of Recyclables. Provide a tenant-accessible area for the collection and storage of materials including glass, paper, metals, cardboard, and plastic.

Credits

- Long-Term Commitment. 1 point; requires that the tenant remain in the same location for no fewer than 10 years to encourage choices that will conserve resources.
- Building Reuse, Maintain 40 or 60% of Interior Nonstructural Components. 2 possible credits. For 1 credit, maintain at least 40% by area of the existing nonshell, nonstructure components, such as walls, floors, and ceilings. Make it 2 credits if it is 60% reuse.
- Construction Waste Management. Up to 2 credits available. Divert from landfills construction-related materials through the use of a waste management plan. Divert 50% for 1 credit or 75% for 2 credits.
- Resource Reuse. 2 possible credits; requires the use of at least 5% of building construction materials, excluding furniture and furnishings for 1 credit. If 10% is reused, it is 2 credits.
- Resource Reuse—30% Furniture and Furnishings. 1 point; requires the use of salvaged, refurbished, or used furniture and furnishings for 30% of the total budget for these items.
- Recycled Content. 2 points possible. The intent is to increase demand for recycled content and materials. These include post- and preconsumer materials used in furniture and furnishings. If 10%, it is 1 credit; if 20%, it is 2 credits.
- Regional Materials 20% Manufactured Regionally. 1 point available. Use a minimum of 20% combined value of furniture, materials, and products that are manufactured regionally within a radius of 500 mi.
- Regional Materials 10% Extracted and Manufactured Regionally. 1 point. In addition to the regional materials above, use a minimum of 10% of the combined value of construction materials and products extracted, harvested, or recovered as well as manufactured within 500 mi.
- Rapidly Renewable Materials. 1 point. Use rapidly renewable materials for 5% of the total value of all materials and products used.
- Certified Wood. 1 point. When using wood products, use a minimum of 50% certified by the Forest Stewardship Council.

Indoor Environmental Quality (IEQ). Up to 17 possible credits in this category. Note the use of the phrase IEQ instead of the more common phrase indoor air quality. IAQ is a fundamental part of IEQ, but IEQ encompasses much more than air quality, as outlined below.

Prerequisites

- Minimum IAQ Performance. Satisfy the requirements for ventilation based upon ASHRAE Standard 62.1-2004, Ventilation for Acceptable Indoor Air Quality.
- Environmental Tobacco Smoke (ETS) Control. The most common path to compliance is simply to prohibit smoking in the building. This has become common practice and is frequently required by codes. Alternative paths to compliance include the construction of dedicated smoking rooms with dedicated HVAC and exhaust systems.

Credits

- Outdoor Air Delivery Monitoring. 1 point; provide permanent monitoring and alarm systems that provide feedback on ventilation system performance.
- Increased Ventilation. 1 point; increase the outdoor air ventilation by 30%, above required by ASHRAE Standard 62.1-2004. However, this can significantly increase the building's energy consumption. Reference the ASHRAE Green Guide for ways to ameliorate the negative energy impacts of this measure.
- Construction IAQ Management Plan. 2 points available. The first point is earned during construction by developing and implementing an IAQ Construction Management Plan. It includes protection from moisture for building materials, special air filters for HVAC during construction, and replacing all filter media prior to occupancy. The second point is for Before Occupancy. It includes optional requirements for building flushout or an IAQ test procedure to ensure suitability of the space before occupancy.
- Low Emitting Materials. Up to 5 points available for five categories of products. The intent is to specify low Volatile Organic Compound (VOC)-emitting products.

Inv–Light

These product categories are adhesives and sealants, paints and coatings, carpet systems, composite wood and laminate adhesives, and systems furniture and seating.

- Indoor Chemical and Pollutant Control. 1 point; minimize the exposure of occupants to potential contaminates and chemical pollutants that can adversely affect IAQ. Methods to achieve this include entryway systems such as grates to trap dirt. Where hazardous products may be used, such as janitorial closets, provide dedicated exhausts and spill containment.
- Controllability of Systems, Lighting. 1 credit; provide lighting controls for at least 90% of the occupants.
- Controllability of Systems, Temperature, and Ventilation. 1 point; provide at least 50% of the occupants the ability to adjust thermal and ventilation conditions to suit individual needs and preferences. This may be an energy-waster, depending upon the HVAC systems in place. Reference the ASHRAE Green Guide for ways to deal with this issue.
- Thermal Comfort Compliance. 1 credit; comply with ASHRAE Standard 55-2004, Thermal Conditions for Human Occupancy. Note that Standard 55 is based upon providing conditions that will satisfy 80% of the population, which implies that 20% may not be comfortable. This may become an energy-efficiency and a comfort dilemma for operations personnel.
- Thermal Comfort, Monitoring. 1 point; provide a permanent monitoring system and process to ensure compliance with the ASHRAE 55 credit listed above.
- Daylight and Views, Daylight for 75 and 90%. Up to 2 credits available; provide occupants a connection between indoor and outdoor spaces through the use of daylight. If a minimum daylight factor of 2% is achieved for 75% of the occupants, it is 1 point. If that factor is for 90% or more of the occupants, it is 2 points.
- Daylight and Views, Views for 90% of the Seated Spaces. 1 point. Achieve direct line of sight to the outdoor environment for at least 90% of those spaces normally occupied.

Innovation and Design Process. Up to 4 points available. This is a way to encourage the innovation and evolution of LEED through the recognition of exceptional performance above the requirements set by LEED. It includes identification of the intent of the credit, the proposed requirements, the proposed submittals, and the design approach.

Leadership in Energy and Environmental Design for Accredited Professional. 1 credit. This is the use of a LEED Accredited Professional (AP) as one of the principal participants in the project team.

LEED Core and Shell, C&S

Based upon the LEED-NC rating system for NC and major renovation, LEED-CS was developed in recognition of the unique nature of C&S developments. In particular, there is the lack of developer control over key aspects such as interior finishes, lighting, and HVAC distribution. Thus, the scope of CS is limited to those elements of the project under the direct control of the developer.

When CS is combined with its companion rating system, LEED-CI for commercial interiors, the USGBC addresses the entire building, core, shell, and interiors, but with responsibility assigned to those parties having direct control of their particular sections. Currently in the ballot phase, CS is expected to be adopted in the fall of 2006.

Precertification

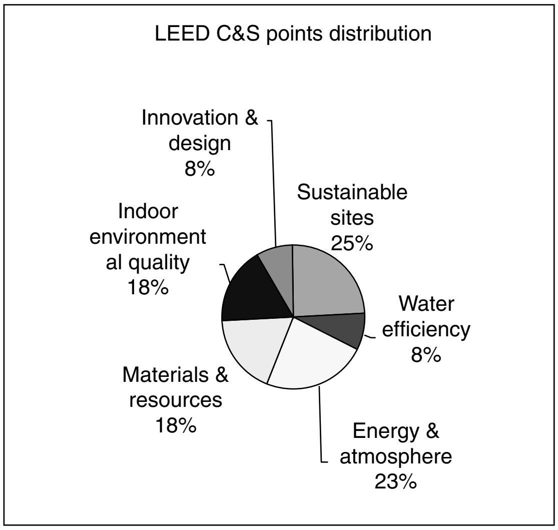
Many times, an LEED-CS building will be built speculatively, without tenants or an interior fitout design. Therefore, the USGBC recognized the need for LEED-CS projects to be precertified prior to the building's construction. This way, the developer can promote his building as a LEED-compliant design and position his building in the competitive marketplace. This is different from other LEED products, wherein certification is awarded after construction is complete. But in fact, what is precertified is only the building's design. Precertification does not guarantee future certification. All it provides is a conditional promise in writing from the USGBC that if in fact the C&S are constructed in accordance with the precertification documents, and if the USGBC full-documentation process is followed, the building will be certified.

LEED-CS Credit Categories

Below is a summary of where the points will be for LEED-CS. Note the similarities to LEED-NC and that Energy and Atmosphere is the largest point category. This particular version of C&S is the ballot version for adoption. Earlier versions varied slightly during the public comment stages, and this ballot represents the final version, pending some last-minute changes.

Possible points	
Sustainable sites	15
Water efficiency	5
Energy and atmosphere	14
Materials and resources	11
Indoor environmental quality	11
Innovation and LEED AP	5
Total points available	61

Inv–Light



Four levels of certification	
Certified	23–27 points
Silver	28–33 points
Gold	34–44 points
Platinum	45–61 points

LEED Core and Shell, C&S, Technical Review

The format of LEED C&S is similar to LEED-NC and CI. There are prerequisites and credits. There are the five basic categories of credits: Sustainable Sites, Water Efficiency, Energy and Atmosphere, Materials and Resources, and Indoor Environmental Quality. There are also credits for Innovation and Design Process.

Sustainable Sites. Maximum available points is 15.

Prerequisites. There is one prerequisite.

- Construction Activity Pollution Prevention. Create and implement an erosion and sedimentation control plan for all construction activities of the project.

Credits

- Site Selection. 1 credit available. The intent is to discourage inappropriate development and to reduce the environmental impacts of the building's location. A list of prohibited development is provided, which includes wetlands, prime farmlands, land especially identified as habitat for threatened or endangered species, and public parkland.
- Development Density and Community Connectivity. 1 point available; encourage development in areas with existing infrastructure. There are two options: Development Density, which is to build on a previously used site and in a community with a minimum density of 60,000 ft^2 per acre, and Community Connectivity, which is to construct on a previously developed site within one-half mile of a residential zone with an average density of 10 units per acre and within one-half mile of basic services with pedestrian access.
- Brownfield Redevelopment. 1 point available; develop on a site documented as contaminated by means of American Society for Testing and Materials (ASTM) E1903-97 Phase II Environmental Site Assessment.
- Alternative Transportation. Up to 4 credits available; 1 credit each for access to public transportation, provision of bicycle storage and changing rooms, low-emitting and fuel-efficient vehicles, parking capacity, and a plan to promote vans and carpooling and discourage single vehicle occupancy.
- Site Development. 2 credits available. The first credit is for protecting or restoring habitat for greenfield sites and limiting site disturbance to 40 ft beyond the building perimeter or, for previously developed sites, restoring or protecting 50% of the site area. The second credit is for maximizing open space. Three options are available. The first option is to reduce the development footprint. The second option, for areas with no local zoning requirements, is to provide vegetated open spaces adjacent to the building. The third option, where a zoning ordinance exists but has no requirement for open space, is to provide vegetated open space equal to 20% of the project site area.
- Stormwater Design. 2 credits available. The first credit is for quantity control of stormwater peak discharges and a management plan to protect stream channels. The second credit refers to quality control of discharges. It refers to an implementation plan to promote infiltration and to capture and treat stormwater runoff from 90% of the annual rainfall.
- Heat Island. 2 credits available. The first credit is for nonroof applications and includes strategies to provide vegetation for shade, high-reflectance paving materials, or open grid pavement, or placing a minimum of 50% of parking spaces under cover. The second credit for roof applications includes strategies that can be roofs with a high Solar Reflectance Index, a green vegetation-covered roof, or a combination of the two.
- Light Pollution Reduction. 1 credit available; reduce light trespass from the building, including outdoor lighting as well as overflow from interior lighting.
- Tenant Design and Construction Guidelines. Provide tenants design and construction information that includes a description of sustainable design features incorporated into the C & S part of the project. This serves to educate and encourage tenants to complete the sustainability promise of the project. It also includes LEED CI guidance.

Inv–Light

Water Efficiency. No prerequisites are required in this section. A total of 5 credits is available in this section.

Credits

- Water Efficient Landscaping. Up to 2 credits available. The first credit is for a reduction of 50% or more in the use of potable water for landscaping, calculated from the midsummer baseline consumption. The second credit is for the use of nonpotable water for irrigation or the use of no irrigation through the careful selection of landscaping vegetation.
- Innovative Wastewater Technologies. 1 credit. Two options are available. The first option is to reduce potable-water use for sewage conveyance by the use of water-conserving plumbing or the use of nonpotable water for flushing. The second option is to treat 50% or more of wastewater on site to tertiary standards and reuse it on site.
- Water-Use Reduction. Up to 2 credits; employ strategies that in aggregate use less than the baseline consumption, as calculated to meet the Energy Policy Act of 1992 performance. A 20% reduction gets 1 credit; a 30% reduction gets 2 credits.

Energy and Atmosphere. Up to 14 credits available in this category. There are three prerequisites required.

Prerequisites

- Fundamental Commissioning of the Building Energy Systems. This refers to a commissioning of the basic energy-consuming systems in the C&S. They include HVAC and R, the building automation system, lighting, lighting controls, domestic hot water systems, renewable energy systems, and daylighting.
- Minimum Energy Performance. The project is to be designed to comply with ASHRAE/IESNA Standard 90.1-2004. This includes the mandatory provisions and either of the following two: the prescriptive requirements, which follow the prescriptive information provided in the standard, or the performance requirements, which are based upon energy-consumption computer modeling of the building's energy systems. The modeling method allows the design team greater flexibility in overall design.
- Fundamental Refrigerant Management. Requires zero use of CFC refrigerants to reduce ozone depletion. In reusing an existing building that contains CFCs, a phaseout plan must be implemented before project completion, although application for special consideration to extend the phaseout will be considered on its merits.

Credits

- Optimize Energy Efficiency. Up to 8 credits available. This is the single largest category for LEED credits. There are three optional paths. Option 1 is to perform a building energy-consumption simulation and compare this with the baseline established as part of the prerequisite for minimum energy performance. Credits are awarded for the percentage that is better than the baseline. For example, 1 credit is given for NC if it is 10.5% better or for an existing building renovation if it is 3.5% better. This goes up to 8 credits for NC when it is 35% better or existing buildings when they are 28% better. Option 2 is a prescriptive path for 3 credits by following the ASHRAE Advanced Energy Design Guide for Small Office Buildings. An additional 3 credits are available in Option 2. One credit is given by improving the building envelope beyond the ASHRAE Design Guide[5] baseline design. Another credit is given for lighting systems that comply with all the ASHRAE Design Guide recommendations, and the third credit is given for HVAC and service water heating systems that comply with all recommendations of the ASHRAE Design Guide. Option 3 can earn only 1 credit, to comply with the Basic Criteria and Prescriptive Measures of the NBI Advanced Buildings Benchmark version 1.1, with the exception of several sections, including Design, Construction, and Operations Certification, Energy Code Compliance, Indoor Air Quality, Refrigeration Equipment Efficiency, Monitoring and Trend Logging, Networked Computer Monitor Control, and Cool Roofs.
- On-Site Renewable Energy. 1 credit; requires that at least 1% of the building's baseline annual energy cost be renewable energy generated on site.
- Enhanced Commissioning. 1 credit. This refers to commissioning more comprehensive than the Fundamental Buildings Commissioning Prerequisite. It includes a design review prior to construction, the review of all submittals for energy-consuming systems, the development of a systems manual for future operations staff, verification that operating personnel have been properly trained, and continued project involvement for at least 10 months after completion by reviewing building operations and developing a plan of resolution for outstanding commissioning-related activities.
- Enhanced Refrigerant Management. 1 credit. There are two optional ways to comply. Option 1 is to not use refrigerants. This may be satisfied by having no mechanical cooling. Option 2 is to select refrigerants for HVAC and R that minimize or eliminate the emission of compounds that contribute to ozone depletion and global warming.
- Measure and Verification (M&V). There are two credits available. The first credit is for the base building, which includes a Measurement & Verification, M & V, Plan that will provide the necessary

infrastructure, usually part of the Building Automation System, BAS. This is to measure base building energy consumption for electricity and other services for least one year of postconstruction occupancy. The second credit is available for tenant submetering to provide ongoing tenant responsibility for energy bills. Again, this can be done through sensors that are part of the BAS.

- Green Power. 1 credit; requires that at least 35% of C&S base-building electricity comes from renewable energy sources by engaging in a green power purchase contract for at least a 2-year duration.

Materials and Resources. Up to 11 credits available, with one prerequisite.

Prerequisite

- Storage and Collection of Recyclables. Provide an easily accessible area that serves the entire building and that is dedicated to the collection and storage of recyclable materials.

Credits

- Building Reuse: Maintain a Percentage of Existing Walls, Floors, and Roof. Up to 3 credits available; extend the life cycle of existing buildings by maintaining a percentage of the existing structures, including walls, floors, and roof. The first credit is for 25% reuse; 50% reuse earns a second credit; and at 75% reuse, a third credit is earned.
- Construction Waste Management. Up to 2 credits available; divert construction and demolition materials from disposal by recycling and reusing a percentage of nonhazardous construction and demolition debris. Develop and implement a plan to identify, sort, and set aside materials for reuse. The first credit is earned for 50% diversion from disposal. An additional credit is earned if 75% is diverted.
- Materials Reuse. Earns 1 credit; use salvaged, refurbished, or reused materials, the sum of which is 1%, based upon the cost of the total value of materials for the project.
- Recycled Content. Up to 2 credits available; requires the use of recycled content such that the sum of postconsumer recycled content plus one-half of the preconsumer constitutes a percentage of the total cost of materials on the project. If that percentage is 10%, 1 credit is earned. For 20% or more, 2 credits are earned.
- Regional Materials: Extracted, Processed and Manufactured Locally. Up to 2 credits available. The intent is to use materials or products that have been extracted, harvested, recovered, or manufactured locally within 500 mi of the job site. If the total is 10% by cost of total materials used, this earns 1 credit. If the total is 20% or more, 2 credits are earned.
- Certified Wood. 1 credit. For all wood products used, ensure that a minimum of 50% of wood-based products is certified in accordance with the Forrest Stewardship Council. This can include furniture as well as building structural components.

Indoor Environmental Quality. There are two prerequisites and a total of 11 credits available in this category.

Prerequisites

- Minimum IAQ Performance. Comply with Sections 4–7 of ASHRAE 62.1-2004, Ventilation for Acceptable Indoor Air Quality.
- Environmental Tobacco Smoke (ETS) Control. Three optional paths for compliance. The first option is to prohibit smoking in the building and locate any designated outdoor smoking areas at least 25 ft away from the building and all air intakes. Prohibition is the simplest option to use. Option 2 consists of several parts: (1) Prohibit smoking except in designated areas; (2) designated smoking areas must have isolated HVAC and exhaust systems, separate from the house systems; (3) use of differential air pressures is required to ensure that smoke does not infiltrate into nonsmoking space; and (4) outdoor designated spaces are to be at least 25 ft away from the building and all building air intakes. The third option is for residential applications. This requires prohibition of smoking in all common areas and locates outdoor smoking areas at least 25 ft away.

Credits

- Outdoor Air Delivery Monitoring. Earns 1 credit; requires permanent monitoring systems to provide feedback on ventilation system performance to ensure that it is working up to design performance. For mechanically ventilated spaces, refer to ASHRAE 62.1-2004. For naturally ventilated spaces, use CO2 monitoring to ensure that ventilation is adequate.
- Increased Ventilation. Earns 1 credit. For mechanically ventilated spaces, increase breathing zone ventilation rates at least 30% more than the minimum required by ASHRAE 62.1-2004. Note that this can come at the cost of large energy losses unless energy-conserving measures such as heat recovery are used. For naturally ventilated spaces, follow the guidelines outlined in the Carbon Trust Good Practice Guide 237 and the Chartered Institution of Building Services Engineers Applications Manual 10:2005.
- Construction IAQ Management Plan: During Construction. 1 credit. Implement an IAQ management plan for the construction preoccupancy phases of the building. This includes meeting the control measures of the Sheet Metal and Air Conditioning National Contractors Association (SMACNA) IAQ Guidelines for Occupied Buildings under Construction, 1995, Chapter 3.[6] This credit also requires protection of construction materials from moisture damage and mold growth.

- Low-Emitting Materials. Up to 3 points available. Four categories of low-emitting materials are listed here. The intent is to reduce the indoor air contaminants that may be odorous, irritating, or harmful—essentially striving to have an indoor environment with a minimum of VOCs and other irritating substances. Categories are adhesives and sealants, paints and coatings, carpet systems, composite wood, and agrifiber products. Compliance with two of these categories earns 1 credit. Compliance with three categories earns 2 credits, and compliance with all four categproes earns 3 credits.
- Indoor Chemical and Pollutant Source Control. Earns 1 credit. Design to control and minimize pollutant entry into buildings. This is accomplished by using entryway systems to prevent dirt and particulates from entering the building. Where hazardous materials may be stored, such as in janitorial closets, provide exhaust systems. And for mechanically ventilated spaces, provide air filtration that has a Minimum Efficiency Ratings Value (MERV) of 13 or better.
- Controllability of Systems. 1 credit. Provide individual comfort controls for a minimum of 50% of the building occupants to enable adjustments to suit individuals' needs. Operable windows can be used in lieu of comfort controls in certain settings. Obviously, this may have a very negative impact on overall building energy consumption unless it is well designed and managed.
- Thermal Comfort: Design.1 credit. Design the HVAC systems and building envelope to comply with ASHRAE Standard 55-2004 for Thermal Comfort. This applies only to C&S projects that are providing the base-building HVAC system. If the tenant is to provide the HVAC, this credit is not available to the C&S applicant. However, it would be available for the tenant under CI. Note that this standard recognizes that 80% of building occupants will find this condition fully acceptable, which implies that possibly 20% will be uncomfortable to some degree.
- Daylighting and Views. Up to 2 points available, depending upon the percentage of occupants who have daylighting as defined in the credit requirements. Daylighting compliance requires a calculation that includes glazing factors, window areas, floor space geometry, and other factors. Details are supplied in the USGBC LEED documents. The intent of these credits is to provide occupants a connection to outdoor spaces through the introduction of daylight and outdoor views. For 75% compliance, the project earns 1 credit. For 90% compliance, it earns 2 credits.
- Innovation and Design Process. Earns up to 4 credits, in similar fashion as described in LEED-CI previously.
- LEED Accredited Professional. Earns 1 credit to have an LEED AP as a principal member of the project team.

You can download all four of the LEED rating systems. The rating-systems downloads are free. However, other tools and workbooks, such as the reference guides, do have a fee associated with them, with discounts given to members:

1. Visit the U.S. Green Building Council Web site at www.usgbc.org/leed.
2. Choose a rating system.
3. Click the rating system you would like to download.

DISCUSSION OF LEED

The USGBC was formed to transform the built environment and the processes that constitute the design, construction, and operation of buildings. However, there are critics who say the USGBC has not lived up to its promise and has not made a difference in the marketplace. Leadership in Energy and Environmental Design and the USGBC have experienced their share of criticism because of this. Typically, these criticisms focus on the complexity of the LEED certification process. It is detail oriented, very demanding, and time consuming, and there is no guarantee that a building will pass certification even after all the work is done. In addition to process concerns, green buildings have been criticized as being too expensive to build. Many in the buildings industry believe that building green might be analogous to "gold plating." Green seems to add much cost with insufficient value added.

Another question has been about the ability of the design community and product suppliers to deliver green products. Are there enough trained professionals? Also, are there enough green products available at comparable costs to standard materials? In addition, some may say that the USGBC is too internally focused and not inclined "to play well with others," such as with other organizations and trade associations.

The USGBC has recognized these problems and concerns, and has acted vigorously to resolve these issues. First, regarding market penetration, although there are still not many certified green buildings, the marketplace is abuzz with arguments—pro and con—about sustainability and green buildings. So although actual penetration is slight when compared with all buildings, almost none of those new buildings is built without a discussion of green. The USGBC has in fact already transformed the marketplace; it has made sustainable green buildings a major topic of discussion and argument. As for interest in, if not outright acceptance of, LEED, it is evident everywhere. It ranges from major corporations to all levels of government and to educational and health-care industries, etc. At this point, the U.S. Army, Air Force, and Navy, and other Department of Defense agencies require LEED.[7]

Rick Fedrizzi, current head of the USGBC and one of its founders, likens LEED to an adolescent. Leadership in Energy and Environmental Design is still growing, maturing, and finding better ways to do its job.[8] Fedrizzi has compared the adaptability of LEED with the U.S. Constitution. Although the authors of the original LEED

NC 2.0 made the best possible document at the time, they recognized that changing forces in the global effort for sustainability would require the evolution of LEED. Thus, LEED has incorporated the innovation credits, which encourage and reward creativity, risk-taking, and innovation. Leadership in Energy and Environmental Design can be amended in a fashion similar to a constitutional amendment. All stakeholders, essentially all USGBC members, have not only the opportunity, but also the responsibility to vote on changes. The current results of this evolution have been the development of new LEED products such as C&S, CI, and Existing Buildings. In the wings are LEED for Homes and Neighbourhood Development. Besides new LEED products, there are variations in the existing programs being developed. There are versions of LEED-NC for schools, labs, health-care facilities, etc. being considered. Another modification to LEED being considered is the regionalization of LEED. Currently, regional advisory teams are being recruited from the membership, through the USGBC local chapters. These teams will consider creating regionalized LEED criteria based upon local factors affecting sustainability. For example, for the arid Southwest, water is a critical resource, and its conservation should have a greater emphasis than it does in the current LEED programs. So maybe additional water-related credit categories will be developed for the Southwest. Similarly, in the mid-Atlantic states, water is less an issue than it is in the Southwest, but open spaces, landfills, and uncontrolled development are greater concerns.

The USGBC has also streamlined the certification process and provided online document templates to facilitate the certification process. Additional efforts to facilitate the process continue.

As for professional development, the USGBC recognized very early that sustainable buildings will need not only a roadmap for sustainable buildings (for example, LEED), but also the professionals to deliver this product. Professional training, seminars, and accreditation have been cornerstones of the USGBC effort, and these educational programs are expanding. The USGBC is determined to ensure that there are enough trained professionals. With these trained professionals comes the development of experienced LEED design teams. In the beginning of sustainable design and construction, green buildings were difficult and more expensive to build than conventional buildings. But now, with experienced teams, newer green projects are coming in at very comparable prices to conventional buildings. The perceived cost premium for green has faded away.

Last, the USGBC has welcomed the membership of trade associations, which were previously prohibited. In addition, other organizations have actively joined with the USGBC. In 2002, ASHRAE signed a memorandum of understanding (MOU) with the USGBC to develop joint programs. Among the results of this teaming have been the development of the ASHRAE Green Guide, the Small Buildings Design Guide and USGBC, ASHRAE/IESNA, and ANSI Standard 189, which will provide code language for the design and construction of green buildings.[9]

CONCLUSION

Where is LEED going? What is the future? We can see there are new products being developed for different markets. We see the regionalization of LEED credits. There is teaming with other interested and related organizations. But the current focus of efforts has been primarily on the design and construction of new and renovated buildings. The focus has been on the architects to lead the sustainable design team, consisting of engineers, interior designers, suppliers, and contractors. But when the new building is constructed, the ribbon at the entranceway has been cut, and the LEED certification plaque is mounted on the wall, everyone shakes hands, and compliments go all around for a job well done! But is it? Is this the end? Or is it the end of the beginning? I would call this a transition from beginning to operations.

Leadership in Energy and Environmental Design certification is similar to the educational process. We send students to the best schools, and they receive the best grades and earn their diplomas. But does it all stop there? No. The graduates now must deliver on their promise of excellence. Over their working life of 30 or 40 years, they must deliver on their promise. It is the same with a building. It will have to be operated in a superior fashion over many years for it also to deliver on that high-performance promise made so many years ago when that ribbon was cut. But then, we do have LEED for Existing Buildings, and that's for another entry.

REFERENCES

1. U.S. Green Building Council (www.USGBC.ORG). *LEED Green Building Rating Systems for Commercial Interiors and Core&Shell*.
2. American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE, (www.ASHRAE.ORG) *Standard 62.1-2004 Ventilation for Acceptable Indoor Air Quality*.
3. ASHRAE Standard 55-2004, *Thermal Environmental Conditions for Human Occupancy*.
4. ASHRAE Standard 90.1-2004, *Energy Standard for Buildings Except Low-Rise Residential Buildings*.
5. ASHRAE, *Advanced Energy Design Guide for Small Office Buildings*.
6. SMACNA, *IAQ Guidelines for Occupied Buildings Under Construction*, (www.SMACNA.org/bookstore/index.cfm).
7. Army Embraces LEED Standard, *FacilitiesNet*, April 18, 2006, www.facilitiesnet.com/news/article2.asp?id=4400.
8. Bowen, T.S.; Constructive Criticism for Green Standards, *GRIST Magazine*, November 4, 2005.
9. ASHRAE Standard 189, *Proposed Standard to Provide Minimum Design Requirements for Sustainable Buildings*.

Inv–Light

LEED-NC: Leadership in Energy and Environmental Design for New Construction

Stephen A. Roosa
Energy Systems Group, Inc., Louisville, Kentucky, U.S.A.

Abstract

The Leadership in Energy and Environmental Design (LEED™) Green Building Rating System is a set of rating systems for various types of construction projects. Developed by the U.S. Green Building Council (USGBC), the rating systems evolved with the intent of helping to "fulfill the building industry's vision for its own transformation to green building." [U.S. Green Building Council. *LEED green building rating system*. 2004.] The rating systems developed by the USGBC are for new construction, existing buildings, core and shell construction, commercial interiors, homes, and residences. This article considers one of these systems, LEED-NC. It discuses the influences that shaped development of LEED-NC for new construction and major renovations, details how the process works, and considers how new projects are scored. The article concludes by providing a brief assessment of the program's strengths and weaknesses.

INTRODUCTION

Land development practices have yielded adverse environmental consequences, urban dislocation, and changes in urban infrastructure. Urban development in particular has long been associated with reduced environmental quality and environmental degradation.[2] The rate at which undeveloped land is being consumed for new structures—and the growing appetite of those structures for energy and environmental resources—has contributed to ecosystem disruption and has fostered impetus to rethink how buildings are sited and constructed. While urban developmental patterns have been associated with environmental disruptions at the local and regional scales, the scientific assessments of global impacts have yielded mixed results. In part as a reaction to U.S. development patterns that have traditionally fostered suburbanization and subsidized automobile-biased transportation infrastructure, design alternatives for structures with environmentally friendly and energy efficient attributes have become available.

According to the United Nations Commission on Sustainable Development, "air and water pollution in urban areas are associated with excess morbidity and mortality … Environmental pollution as a result of energy production, transportation, industry or lifestyle choices adversely affects health. This would include such factors as ambient and indoor air pollution, water pollution, inadequate waste management, noise, pesticides and radiation."[3] It has been demonstrated that a relationship exists between the rates at which certain types of energy policies are adopted at the local level and select indicators of local sustainability.[4] As more urban policies focus on the built environment, buildings continue to be the primary building blocks of urban infrastructure. If buildings can be constructed in a manner that is less environmentally damaging and more energy efficient, then there is greater justification to label them as "green" buildings.

The concept of sustainability has evolved from considerations of land development, population growth, fossil fuel usage, pollution, global warming, availability of water supplies, and the rates of resource use.[5] Thankfully, a vocabulary of technologies and methodologies began to develop in the 1970s and 1980s that responded to such concerns. Driven by ever increasing energy costs, energy engineers began to apply innovative solutions, such as use of alternative energy, more efficient lighting systems and improved electrical motors. Controls engineers developed highly sophisticated digital control systems for heating, ventilating and air conditioning systems. With growing concerns about product safety and liability issues regarding the chemical composition of materials, manufacturers began to mitigate the potential adverse impacts of these materials upon their consumers. Resource availability and waste reduction became issues that began to influence product design. In the span of only 25 years, local governments made curbside recycling programs in larger U.S. cities nearly ubiquitous. Terms and phrases such as "mixed use planning," "brownfield redevelopment," "alternative energy," "micro-climate," "systems approach," "urban heat island effect," "energy assessments," "measurement and verification," and "carrying capacity" created the basis for a new vocabulary

Keywords: Leadership in energy and environmental design (LEED); Sustainability; Sustainable development; Planning; Indoor air quality (IAQ); LEED for new construction.

which identifies potential solutions. All of these concerns evolved prior to the 1992 U.N. Conference on the Environment and Development, which resulted in the Rio Agenda 21 and clarified the concept sustainability.

In regard to the built environment, architectural designers renewed their emphasis on fundamental design issues, including site orientation, day lighting, shading, landscaping, and more thermally cohesive building shells. Notions of "sick building syndrome" and illnesses like Legionnaires' disease, asthma and asbestosis, jolted architects and engineers into re-establishing the importance of the indoor environmental conditions in general and indoor air quality (IAQ) in particular when designing their buildings.

The decisions as to what sort of buildings to construct and what construction standards to apply are typically made locally. Those in the position to influence decisions in regard to the physical form of a proposed structure include the builder, developer, contractors, architects, engineers, planners, and local zoning agencies. In addition, all involved must abide by regulations that apply to the site and structure being planned. The rule structure may vary from one locale to another. What is alarming is that past professional practice within the U.S. building industry has only rarely gauged the environmental or energy impact of a structure prior to its construction. Prior to the efforts of organizations like the U.S. Green Building Council (USGBC) (established in 1995), the concept of what constituted a "green building" in the United States lacked a credible set of standards.

THE CONCEPT OF GREEN BUILDINGS

Accepting the notion that sustainable, environmentally appropriate, and energy efficient buildings can be labeled "green," the degree of "greenness" is subject to multiple interpretations. The process of determining which attributes of a structure can be considered "green" or "not green" is inconclusive and subjective. Complicating the process, there are no clearly labeled "red" edifices with diametrically opposing attributes. While it is implied that a green building may be an improvement over current construction practice, the basis of attribute comparison is often unclear, subjective, and confusing. It is often unclear as to what sort of changes in construction practice, if imposed, would lead the way to greener, more sustainable buildings. If determinable, the marketplace must adjust and provide the technologies and means by which materials, components, and products can be provided to construction sites where greener buildings can arise. Since standards are often formative and evolving, gauging the degree of greenness risks the need to quantify subjective concepts.

There are qualities of structures, such as reduced environmental impact and comparatively lower energy usage, which are widely accepted as qualities of green construction practices. For example, use of recycled materials with post-consumer content that originates from a previous use in the consumer market and post-industrial content that would otherwise be diverted to landfills is widely considered an issue addressable by green construction practices. However, evaluation of green building attributes or standards by organizations implies the requirement that decisions be based on stakeholder consensus. This process involves input to the decision-making processes by an array of representative stakeholders in often widely diverse geographic locations. For these and other reasons, developing a rating system for green buildings is both difficult and challenging.

RATING SYSTEMS FOR BUILDINGS

Rating systems for buildings with sustainable features began to emerge in embryonic form in the 1990s. The most publicized appeared in the United Kingdom, Canada, and the United States. In the United Kingdom, the Building Research Establishment Environmental Assessment Method (BREEAM) was initiated in 1990. BREEAM™ certificates are awarded to developers based on an assessment of performance in regard to climate change, use of resources, impacts on human beings, ecological impact, and management of construction. Credits are assigned based on these and other factors. Overall ratings are assessed according to grades that range from pass to excellent.[6]

The International Initiative for a Sustainable Built Environment, based in Ottawa, Canada, has its Green Building Challenge program with more than 15 countries participating. The collaborative venture is geared toward the creation of an information exchange for sustainable building initiatives and the development of "environmental performance assessment systems for buildings."[7] In the United States, agencies of the central government co-sponsored the development of the Energy Star™ program, which provides "technical information and tools that organizations and consumers need to choose energy-efficient solutions and best management practices."[8] Expanding on their success, Energy Star™ developed a building energy performance rating system which has been used for over 10,000 buildings.

Entering the field at the turn of the new century, the USGBC grew from an organization with just over 200 members in 1999 to 3500 members by 2003.[9] The LEED™ rating system is a consensus-developed and reviewed standard, allowing voluntary participation by diverse groups of stakeholders with interest in the application and use of the standard. According to Boucher, "the value of a sustainable rating system is to condition the marketplace to balance environmental guiding principles and issues, provide a common basis to communicate performance, and to ask the right questions at the start of a project."[10] The first dozen pilot projects using the rating system were certified in 2000.

THE LEED-NC RATING SYSTEM

The USGBC's Green Building Rating System is a voluntary, consensus-developed set of criteria and standards. This rating system evolved with a goal of applying standards and definition to the idea of high-performance buildings. The use of sustainable technologies is firmly established within the LEED project development process. LEED loosely defines green structures as those that are "healthier, more environmentally responsible and more profitable."[1]

LEED-NC 2.1 is the USGBC's current standard for new construction and major renovations. It is used primarily for commercial projects such as office buildings, hotels, schools, and institutions. The rating system is based on an assessment of attributes and an evaluation of the use of applied standards. Projects earn points as attributes are achieved and the requirements of the standards are proven. Depending on the total number of points a building achieves upon review, the building is rated as Certified (26–32 points), Silver (33–38 points), Gold (39–51 points) or Platinum (52 or more points).[11] Theoretically, there are a maximum of 69 achievable points. However, in real world applications, gaining certain credits often hinders the potential of successfully meeting the criteria of others. While achieving the rating of Certified is relatively easily accomplished, obtaining a Gold or Platinum rating is rare and requires both creativity and adherence to a broad range of prescriptive and conformance-based criteria.

The LEED process involves project registration, provision of documentation, interpretations of credits, application for certification, technical review, rating designation, award, and appeal. Depending on variables such as project square footage and USGBC membership status, registration fees can range up to $7500 for the process.[12]

LEED PREREQUISITIES CATEGORIES AND CRITERIA

To apply for the LEED labeling process, there are prerequisite project requirements which earn no points. For example, in the Sustainable Sites category, certain procedures must be followed to reduce erosion and sedimentation. In the category of Energy and Atmosphere, minimal procedures are required for building systems commissioning. Minimal energy performance standards must be achieved (e.g., adherence to ANSI/ASHRAE/IESNA Standard 90.1-1999, Energy Standard for Buildings Except Low-Rise Residential Buildings, or the local energy code if more stringent), and there must be verification that CFC refrigerants will not be used or will be phased out. In addition, there are prerequisite requirements outlining mandates for storage and collection of recyclable material, minimum IAQ performance (the requirements of ASHRAE Standard 62-1999, Ventilation for Acceptable Indoor Air Quality must be adhered to), and the requirement that non-tobacco smokers not be exposed to smoke.

In addition to the prerequisite requirements, the LEED process assigns points upon achieving certain project criteria or complying with certain standards. The total points are summed to achieve the determined rating. Projects can achieve points from initiatives within the following sets of categories: Sustainable Sites (14 points), Water Efficiency (5 points), Energy and Atmosphere (17 points), Materials and Resources (13 points), and Indoor Environmental Quality (15 points). Use of a LEED Accredited Professional (1 point) to assist with the project[13] earns a single point. Additional points are available for Innovation and Design Process (maximum of 4 points).

Within each category, the specific standards and criteria are designed to meet identified goals. In the category of Sustainable Sites, 20.2% of the total possible points are available. This category focuses on various aspects of site selection, site management, transportation and site planning. The goals of this category involve reducing the environmental impacts of construction, protecting certain types of undeveloped lands and habitats, reducing pollution from development, conserving natural areas and resources, reducing the heat island impacts, and minimizing light pollution. Site selection criteria are designed to direct development away from prime farmland, flood plains, habitat for endangered species and public parkland. A development density point is awarded for projects that are essentially multi-story. If the site has documented environmental contamination or is designated by a governmental body as a brownfield, another point is available. In regard to transportation, four points are available for locating sites near publicly available transportation (e.g., bus lines or light rail), providing bicycle storage and changing rooms, provisions for alternatively fueled vehicles and carefully managing on-site parking. Two points in this category are obtained by limiting site disturbances and by exceeding "the local open space zoning requirement for the site by 25%."[14] In addition, points are available by following certain storm water management procedures, increasing soil permeability, and attempting to eliminate storm water contamination. Potential urban heat island effects are addressed by crediting design attributes such as shading, underground parking, reduced impervious surfaces, high albedo materials, reflective roofing materials, or vegetated roofing. Finally, a point is available for eliminating light trespass.

Water efficiency credits comprise 7.2% of the total possible points. With the goal of maximizing the efficiency of water use and reducing the burden on water municipal systems, points are credited for reducing or eliminating potable water use for site irrigation, capturing and using rainwater for irrigation, and using drought tolerant or indigenous landscaping. This section of the LEED standard also addresses a building's internal water consumption. Points are available for lowering aggregate water consumption and reducing potable water use. Reducing the wastewater quantities or providing on-site tertiary wastewater treatment also earns points.

Energy and Atmosphere is the category that offers the greatest number of points, 24.6% of the total possible. The intents of this category include improving the calibration of equipment, reducing energy costs, supporting alternative energy, reducing the use of substances that cause atmospheric damage, and offering measurement and verification criteria. Optimizing the design energy cost of the regulated energy systems can achieve a maximum of ten points. To assess the result, project designs are modeled against a base case solution which lacks certain energy-saving technologies. Interestingly, the unit of measure for evaluating energy performance to achieve credits is not kilocalories or million Btus, but dollars. Points are awarded in whole units as the percentage of calculated dollar savings increases incrementally. In addition to the ten points for energy cost optimization, a maximum of three additional points is available for buildings that use energy from on-site renewable energy generation. Purchased green power is allocated a single point if 50% of the electrical energy (in kWh) comes from a two year green power purchasing arrangement. This category provides points for additional commissioning and elimination of the use of HCFCs and halon gases. Measurement and Verification (M&V) is allowed a point, but only if M&V options B, C, and D, as outlined in the 2001 edition of the International Measurement and Verification Protocol (IPMVP), are used.

The Materials and Resources category represents 18.8% of the total possible points. This category provides credit for material management; adaptive reuse of structures; construction waste management; resource reuse; use of material with recycled content; plus the use of regionally manufactured materials, certain renewable materials and certified wood products. A point is earned for providing a space in the building for storage and collection of recyclable materials such as paper, cardboard, glass, plastics and metals. A maximum of three points is available for the adaptive reuse of existing on-site structures and building stock. The tally increases with the extent to which the existing walls, floor, roof structure, and external shell components are incorporated into the reconstruction. LEED-NC 2.1 addresses concerns about construction waste by offering a point if 50% of construction wastes (by weight or volume) are diverted from landfills and another point if the total diversion of wastes is increased to 75%. A project that is composed of 10% recycled or refurbished building products, materials, and furnishings gains an additional two points. Another two points are available in increments (one point for 5%, two points for 10%) if post-consumer or post-industrial recycled content (by dollar value) is used in the new construction. To reduce environmental impacts from transportation systems, a point is available if 20% of the materials are manufactured regionally (defined as being within 500 miles or roughly 800 km of the site), and an added point is scored if 50% of the materials are extracted regionally. A point is available if rapidly renewable materials (e.g., plants with a ten year harvest cycle) are incorporated into the project, and yet another point is earned if 50% of the wood products are certified by the Forest Stewardship Council.

The category of Indoor Environmental Quality allows 21.7% of the possible total points available. The goals include improving IAQ, improving occupant comfort, and providing views to the outside. With ASHRAE Standard 62-1999 as a prerequisite, an additional point is available for installing CO_2 monitoring devices in accordance with occupancies referenced in ASHRAE Standard 62-2001, Appendix C. A point is also available for implementing technologies that improve upon industry standards for air change effectiveness or that meet certain requirements for natural ventilation. Systems that provide airflow using both underfloor and ceiling plenums are suggested by LEED documentation as a potential ventilation solution. Points are available for developing and implementing IAQ management plans during construction and prior to occupancy. The requirements include using a Minimum Efficiency Reporting Value (MERV) 13 filter media with 100% outside air flush-out prior to occupancy. There are points available for use of materials that reduce the quantity of indoor air pollutants in construction caused by hazardous chemicals and by volatile organic compounds in adhesives, sealants, paints, coatings, composite wood products, and carpeting. A point is offered for provision of perimeter windows and another for individual control of airflow, temperature, and lighting for half of the non-perimeter spaces. Points are available for complying with ASHRAE Standard 55-1992 (Thermal Environmental Conditions for Human Occupancy), Addenda 1995, and installing permanent temperature and humidity control systems. Finally, points are gained for providing 75% of the spaces in the building with some form of daylighting and for providing direct line of-sight vision for 90% of the regularly occupied spaces.

In the category of Innovation and Design Process, 7.2% of the total possible points are available. The innovation credits offer the opportunity for projects to score points as a result of unusually creative design innovations, such as substantially exceeding goals of a given criteria or standard.

ASSESSING LEED-NC

The LEED-NC process has numerous strengths. Perhaps the greatest is its ability to focus the owner and design team on addressing select energy and environmental considerations early in the design process. The LEED design process brings architects, planners, energy engineers, environmental engineers, and IAQ professionals into the program at the early stages of design development. The team adopts a targeted LEED rating as a goal for the project. A strategy evolves based on selected criteria. The team members become focused on fundamental green design practices that have often been overlooked when traditional design development processes were employed.

Furthermore, the LEED program identifies the intents of the environmental initiatives. Program requirements are stated and acceptable strategies are suggested. Scoring categories attempt to directly address certain critical environmental concerns. When appropriate, the LEED-NC program defers to engineering and environmental standards developed outside of the USGBC. The components of the program provide accommodation for local regulations. Case study examples, when available and pertinent, are provided and described in the LEED literature. To expedite the process of documenting requirements, letter templates and calculation procedures are available to program users. The educational aspects of the program, which succinctly describe select environmental concerns, cannot be understated. A Web site provides updated information on the program with clarifications of LEED procedures and practice. The training workshops sponsored by the USGBC are instrumental in engaging professionals with a wide range of capabilities.

These considerations bring a high degree of credibility to the LEED process. Advocates of the LEED rating system have hopes of it becoming the pre-eminent U.S. standard for rating new construction that aspires to achieve a "green" label. To its credit, it is becoming a highly regarded standard and continues to gain prestige. Nick Stecky, a LEED Accredited Professional, firmly believes that the system offers a "measurable, quantifiable way of determining how green a building is."[15]

Despite its strengths, the LEED-NC has observable weaknesses. The LEED-NC registration process can appear to be burdensome, and has been perceived as slowing down the design process and creating added construction cost. Isolated cases support these concerns. Kentucky's first LEED-NC school, seeking a Silver rating, was initially estimated to cost over $200/ft^2 ($2152/m^2) compared to the local standard costs of roughly $120/ft^2 ($1290/m^2) for non-LEED construction. However, there are few comparative studies available to substantially validate claims of statistically significant cost impact. Alternatively, many case studies suggest that there is no cost impact as a result of the LEED certification process. It is also possible that the savings resulting from the use of certain LEED standards (e.g. reduced energy use) can be validated using life-cycle costing procedures. Regardless, LEED-NC fails as a one-size-fits-all rating system. For new construction, Kindergarten to 12th-grade (K-12), school systems in New Jersey, California, and elsewhere have adopted their own sustainable building standards.

There are other valid concerns in regard to the use of LEED-NC. In an era when many standards are under constant review, standards referenced by LEED are at times out of date. The ASHRAE Standard 90.1-1999 (without amendments) is referenced throughout the March 2003 revision of LEED-NC. However, ASHRAE 90.1 was revised, republished in 2001, and the newer version is not used as the referenced standard. Since design energy costs are used to score Energy and Atmosphere points, and energy use comparisons are baselined against similar fuels, cost savings from fuel switching is marginalized. In such cases, the environmental impact of the differential energy use remains unmeasured, since energy units are not the baseline criteria. There is no energy modeling software commercially available that has been specifically designed for assessing LEED buildings. LEED allows most any energy modeling software to be used, and each has its own set of strengths and weaknesses when used for LEED energy modeling purposes. It is possible for projects to comply with only one energy usage prerequisite, applying a standard already widely adopted, and still become LEED certified. In fact, it is not required that engineers have specialized training or certification to perform the energy models. Finally, LEED documentation lacks System International (SI) unit conversions, reducing its applicability and exportability.

A number of the points offered by the rating system are questionable. While indoor environmental quality is touted as a major LEED concern, indoor mold and fungal mitigation practices, among the most pervasive indoor environmental issues, are not addressed and are not necessarily resolvable using the methodologies prescribed. It would seem that having a LEED-accredited professional on the team would be a prerequisite rather than a optional credit. Projects in locations with abundant rainfall or where site irrigation is unnecessary can earn a point by simply documenting a decision not to install irrigation systems. The ability of the point system to apply equally to projects across varied climate classifications and zones is also questionable and unproven.

While an M&V credit is available, there is no requirement that a credentialed measurement and verification professional be part of the M&V plan development or the review process. Without the rigor of M&V, it is not possible to determine whether or not the predictive preconstruction energy modeling was accurate. The lack of mandates to determine whether or not the building actually behaves and performs as intended from an energy cost standpoint is a fundamental weakness. This risks illusionary energy cost savings. Finally, the M&V procedures in the 2001 IPMVP have undergone revision and were not state-of- the-art at the time that LEED-NC was updated in May 2003. For example, there is no longer a need to exclude Option A as an acceptable M&V alternative.

The LEED process is not warranted and does not necessarily guarantee that in the end, the owner will have a "sustainable" building. While LEED standards are more regionalized in locations where local zoning and building laws apply, local regulations can also preempt certain types of green construction criteria. Of greater concern is that it is possible for a LEED certified building to devolve into a building that would lack the qualities of a certifiable building. For example, the owners of a building may choose to remove bicycle racks, refrain from the purchase of green energy after a couple of years, disengage control systems, abandon their M&V program, and remove recycling centers—yet retain the claim of owning a LEED certified building.

CONCLUSION

The ideal of developing sustainable buildings is a response to the environmental impacts of buildings and structures. Developing rating systems for structures is problematic due to the often subjective nature of the concepts involved, the ambiguity or lack of certain standards, and the local aspects of construction. While there are a number of assessment systems for sustainable buildings used throughout the developed world, LEED-NC is becoming a widely adopted program for labeling and rating newly constructed "green" buildings in the United States. Using a point-based rating system, whereby projects are credited for their design attributes, use of energy, environmental criteria, and the application of select standards, projects are rated as Certified, Silver, Gold, or Platinum.

The LEED-NC program has broad applicability in the United States and has been proven successful in rating roughly 150 buildings to date. Its popularity is gaining momentum. Perhaps its greatest strength is its ability to focus the owner and design team on energy and environmental considerations early in the design process. Today, there are over 1700 projects that have applied for LEED certification. Due to the program's success in highlighting the importance of energy and environmental concerns in the design of new structures, it is likely that the program will be further refined and updated in the future to more fully adopt regional design solutions, provide means of incorporating updated standards, and offer programs for maintaining certification criteria. It is likely that the LEED program will further expand, perhaps offering a separate rating program for K-12 educational facilities. Future research will hopefully respond to concerns about potential increased construction costs and actual energy and environmental impacts.

REFERENCES

1. U.S. Green Building Council. *LEED green building rating system*. 2004.
2. Spirn, A.W. *The Granite Garden: Urban Nature and Human Design*; Basic Books: New York, 1984.
3. United Nations. *Indicators of Sustainable Development: Guidelines and Methodologies*; United Nations: New York, 2001; 38.
4. Roosa, S.A. *Energy and Sustainable Development in North American Sunbelt Cities*; RPM Publishing: Louisville, KY, 2004.
5. Koeha, T. What is Sustainability and Why Should I Care? *Proceedings of the 2004 world energy engineering congress*, Austin, TX, Sept 22–24, 2004; AEE: Atlanta 2004.
6. URS. http://www.urseurope.com/services/engineering/engineering-breeam.htm (accessed Feb 2005).
7. iiSBE. International Initiative for a Sustainable Built Environment http://iisbe.org/iisbe/start/iisbe.htm (accessed Feb 2005).
8. United States Environmental Protection Agency. *Join energy star—Improve your energy efficiency*. www.energystar.gov. (accessed June 2003).
9. Gonchar, J. Green building industry grows by leaps and bounds; Engineering News-Record; 2003.
10. Boucher, M. Resource efficient buildings—Balancing the bottom line. *Proceedings of the 2004 world energy engineering congress*, Austin, TX, Sept 22–24, 2004; AEE: Atlanta 2004.
11. U.S. Green Building Council. *Green Building Rating System for New Construction and Major Renovations Version 2.1 Reference Guide*; May 2003; 6.
12. U.S. Green Building Council. *LEED—Certification Process*. www.usgbc.org/LEED/Project/certprocess.asp. (accessed Oct 2004).
13. U.S. Green Building Council. *Green Building Rating System for New Construction and Major Renovations Version 2.1*: Nov 2002.
14. U.S. Green Building Council. *Green Building Rating System for New Construction and Major Renovations Version 2.1*: Nov 2002; 10.
15. Stecky, N. Introduction to the ASHRAE greenguide for LEED, *Proceedings of the 2004 world energy engineering congress*, Austin, TX, Sept 22–24 2004; AEE: Atlanta 2004.

Life Cycle Costing: Electric Power Projects

Ujjwal Bhattacharjee
Energy Engineering Program, University of Massachusetts—Lowell, Lowell, Massachusetts, U.S.A.

Abstract
Life-cycle cost (LCC) assessment involves the estimation of major expected costs within the useful life of a power system. Life-cycle cost estimation facilitates investment decisions before acquiring or developing assets associated with a power project. Life-cycle cost analysis allows comparison of different investment alternatives and thus enables determination of the most cost-effective system. A framework has been developed to conduct LCC analysis of energy projects, with the focus on two phases of a project's life: (i) the development, construction, and commissioning phase and (ii) the operation phase. The energy projects considered are: (i) a 520 MW coal power plant; (ii) a 50 MW wind power project; and (iii) a 3-kWp grid-tied residential PV system. The LCC assessment considers costs associated with the debt servicing, variable costs, interest on working capital, depreciation, annual operation and maintenance costs, etc. The results indicate that the installed cost of wind and coal power projects is comparable. However, the LCC of electricity from the wind power project is around twice that of the coal power plant. The cost of electricity from the PV system is around ten times higher than that from the coal power source.

In this study, a framework has been developed to quantify the financial and fiscal incentives (FIs) offered by federal and state agencies to promote renewable energy technologies. This enables studying the effect of FIs on the LCC of electricity generated from the PV and wind power systems. The state-offered FIs considered are sales tax exemption on the purchase of system components, property tax exemption, income tax rebate, and capital subsidy on the system cost. The FIs offered by the federal agencies are the accelerated depreciation for capital recovery, the production tax incentive, and the income tax rebate. For the residential PV system, the state offered upfront capital subsidy on the system cost and avoided cost of electricity due to onsite power generation contribute in reducing electricity cost (LCC) by 65%. Similarly, the LCC of electricity from the 50 MW wind power project is significantly reduced when the FIs are considered. The accelerated depreciation for capital cost recovery and the production tax credit contribute in reducing the cost of electricity by 65%. The LCC of electricity generated from the PV and wind-based energy systems could be significantly less than that generated from coal power projects when the effects of financial incentives are considered.

INTRODUCTION

Life-cycle cost (LCC) assessment involves the estimation of major expected costs within the useful life of an asset. Life-cycle cost estimation facilitates investment decisions for acquiring or developing an asset. According to Woodward (Ref. 13, p. 336), "LCC of a physical asset begins when its acquisition is first considered and ends when it is finally taken out of service for disposal or redeployment (when a new LCC begins)." LCC analysis allows comparison of different investment alternatives and, thus, the determination of the most cost-effective system. For example, decisions based on a low initial cost do not guarantee low LCC. To conduct LCC, the major costs associated with the development of a project needs to be evaluated. The major costs associated with a project can broadly be categorized into the following:

- Initial cost consisting of
  - — Acquisition and financing costs,
  - — Procurement cost, and
  - — Installation and commissioning costs
- Operation and maintenance costs
- Cost of asset disposal after its useful life

To illustrate the LCC for energy projects, we will consider three case studies. These are: (i) a 520 MW coal power project; (ii) a 50 MW utility scale wind project; and (iii) a 3-kW grid-tied residential solar power unit. The rationale for choosing coal, wind, and solar-based power projects is twofold. First, the relevance of the energy source used for electricity generation is considered and, second, the complexity involved in calculating LCC associated with the scale of a project is reviewed.

The relevance of the three energy sources is as follows: coal has been and is expected to remain a major source of electricity generation. For example, in 1971, coal contributed 40% to the total global electricity supply,

Keywords: Life-cycle cost; Coal; Wind; PV; Financial incentives.

while the corresponding value in 2002 was 39%.[8] Wind has made substantial inroads into the utility scale commercial electricity supply during the last 10–15 years. In recent years, distributed grid-tied residential solar photovoltaic (PV) power systems have been demonstrating viability due to state-(and utility-) offered financial (and fiscal) incentives, especially in Japan, Germany, and several states in the United States.

In terms of the complexities involved in the LCC calculations, those for a small PV power system would be simple in comparison to calculations for a large coal power project. For example, a typical 3-kWp (DC) residential solar system could be procured and installed quickly. On the other hand, a large coal power project (as well as nuclear and large hydro power projects) could take several years to construct and commission.

In this article, major cost components associated with the different power systems are briefly described in "Coal, Wind, and Solar Power Systems." In "Methodology for Calculation of the Life-Cycle Cost," a simple framework for the LCC calculation has been developed. The results of the LCC calculations for the three different power systems are discussed in "Results and Discussion," with concluding remarks in "Conclusions."

COAL, WIND, AND SOLAR POWER SYSTEMS

The three power systems, i.e., coal, wind, and PV power systems, operate in significantly different manners. Therefore, the costs associated with each differ. For example, the initial cost for installing a PV system is high as compared to that for coal and wind power projects. However, the resource needed for electricity generation for the PV is free and the same holds true for a wind power system.

Inv–Light

Coal Power Plant

A coal power plant consists of the steam and turbine generator sections. Major mechanical systems of the plant are: water system, heat cycle make-up system, coal handling plant, ash handling plant, fuel oil system, compressed air system, ventilation and air-conditioning system, fire protection systems, and miscellaneous auxiliaries. The electrical systems are: generator, transformer, switchyard, control and instrumentation systems. Coal power plants are usually constructed either near coal mines (in order to reduce the transportation cost of coal) or near load centers to reduce the electricity transmission losses. Coal power plants are generally operated through onsite "operational and maintenance" personnel. Once the plant has outlived its designed life (around 25–30 years), it is usually retrofitted and made operational.

Wind Power Project

A wind power plant consists of turbines and towers, foundations, access roads, switch gear, instrumentation and control systems, etc. Small wind power projects are generally remotely operated. In general, for every 10–20 turbines, one operator is needed.[11] Wind power projects offer potential socioeconomic benefits to local communities. Typically, for a 50 MW power project, 50 jobs are created during the construction phase. During the operation phase one job is created for every 5–8 MW of the installed capacity.[12] The turbines are repowered, replacing the old and typically smaller wind turbines with new equipment when the design life of the equipment diminishes.[11]

Photovoltaic Power System

A small rooftop (3-kWp) PV system consists of around 20–30 PV modules tied on to panel mountings with the roof. Direct grid-tied inverters convert the DC power produced by the PV to AC power. The safety equipment includes a DC combiner box for wiring, a DC disconnect to facilitate servicing, and an AC disconnect to isolate the PV system from the grid.

Cost Components of Energy Projects

The costs associated with the useful life of a project are categorized (in Introduction) as initial cost, operation and maintenance cost, and disposal cost. Wind and coal power projects usually undergo retrofit after useful design life. Therefore, disposal or decommissioning of these plants is not considered in this study. Similarly, PV modules disposed from the rooftop may not be associated with high costs. Additionally, the recycling of PV modules is unlikely to produce high value products.[6]

The initial costs associated with a power project before commencement of construction activities are typically the project site survey, land lease agreements, power purchase agreements, etc. (see Table 1) for coal and wind power projects. For the PV system, an assessment of energy produced from solar irradiance is completed at the location where the system is to be installed.

During the construction period, the costs are divided into two categories: (i) capital cost and (ii) noncapital cost. Capital costs are associated with the procurement of physical assets, such as boiler, wind turbine, and PV panels, respectively, for coal, wind, and PV power projects. The noncapital costs are associated with services needed for the deployment of physical assets, e.g., costs associated with legal services, erection, testing, and commissioning. In Table 2, the capital and noncapital costs associated with coal, wind, and PV power systems are provided. During the operation phase of the plant, the costs usually considered are variable costs (cost of fuel

Table 1 Activities During Project Development Phase 520-MW coal power project

- Site survey
- Environmental impact assessment
- Fuel and water transportation corridor survey
- Development of project technical specification
- Permitting
- Land lease agreement
- Appointment of engineering procurement and construction contractor
- Debt and equity finance tie-up
- Power purchase agreement
- Fuel supply agreement

50-MW wind power project

- Wind resource assessment and site survey
- Environmental impact assessment (including the bird sighting study)
- Permitting
- Land options agreement
- Power purchase agreement
- Debt and equity finance tie-up

3-kW PV system

- Assessment of energy available at the installed roof tilt angle
- Shading analysis

and consumables), operation and maintenance costs, debt servicing costs, etc. (the environmental costs associated with emissions and solid wastes during the operation of the coal power plant have not been considered in this study).

These are explained in detail in "Methodology for Calculation of the Life-Cycle Cost."

METHODOLOGY FOR CALCULATION OF THE LIFE-CYCLE COST

In this section, a framework for the calculation of LCC has been developed. The LCC considers costs associated with two phases of a power system, i.e., the construction and commissioning phase and the operation phase. Therefore, the costs associated during the former are due to interest during construction (IDC), financing charges, and taxes and duties (Fig. 1). Those associated with the operation phase are due to depreciation, interest on working capital, loan repayment, and variable costs (Fig. 1).

Annual Electricity Generation

The annual electricity generated from the ith power project, E_i (in kWh), is calculated considering the number of hours a plant operates in a year. While annual energy generated is strongly related to the wind speed for a wind power plant, the calorific value of the coal used is important for a coal power plant.

$$E_i = 8760 \times 1000 \times P_{\mathrm{cap}} \times f_{\mathrm{e},i} \tag{1a}$$

where P_{cap} (in MW) is the rated capacity of the plant. $f_{\mathrm{e},i}$ is the effective capacity factor of the ith power plant (i.e., coal or wind power plants). The effective capacity factor considers the plant's overall efficiency of energy conversion and the auxiliary power consumed to operate several electrical and mechanical systems associated with the plant. For example, around 10% of the total electricity produced in a coal power plant is utilized as auxiliary power.[3]

Annual electricity generated from a PV system, E_{PV} (in kWh), is related to annual average daily equivalent sunshine hours [ESH; the ESH is equal to the average monthly daily solar irradiation in kWh/m^2/day. PV modules are rated at 1 kWh/m^2 incident irradiation (or one sun). In other words, ESH represents the capacity factor for a PV system. It indicates the number of hours per day that the system is expected to operate] of the place the PV system is installed.

Table 2 Capital and noncapital costs of energy projects

Capital cost category	Noncapital cost category
520-MW coal power project	
Steam generator	Erection testing and commissioning
Turbine generator	Construction insurance
Balance of plant	Engineering and overheads
Mechanical Systems	Preliminary expenses
Coal handling plant	Land and site development
Ash handling plant	Owner's engineers expenses
Cooling towers and circulating water system	Operators training
Other mechanical systems	Start-up fuel
Electrical Systems	Legal fees, taxes, and duties
Generator and transformer	Establishment expenses
Switchyard	
Balance of electrical systems	
Instrumentation and control systems	
Initial spares	
External water supply system	
Intake channel and pipes	
Related equipment and systems	
Civil works	
Access and diversion roads	
Ash disposal area development	
Coal transportation railway construction	
Residential township	
Temporary constructions	
Plant building facilities	
External coal transportation system	
Merry-go-round rail system	
Rolling stock and locomotive	
Miscellaneous fixed assets	
50-MW wind power project	
Turbine and tower	Erection testing and commissioning
Foundations	Construction insurance
Electrical infrastructure	Engineering and overhead
Substation	Assembly and checkout
Transmission upgrade	Preliminary expenses
Roads and grading	Legal fees, taxes, and duties
Control systems	
Control buildings	
Central building	
3-kW Grid-tied photovoltaic system	
Photovoltaic modules	Assembly and installation
Inverter	
DC and AC disconnects	
Wire and cables	
Panel mounts	

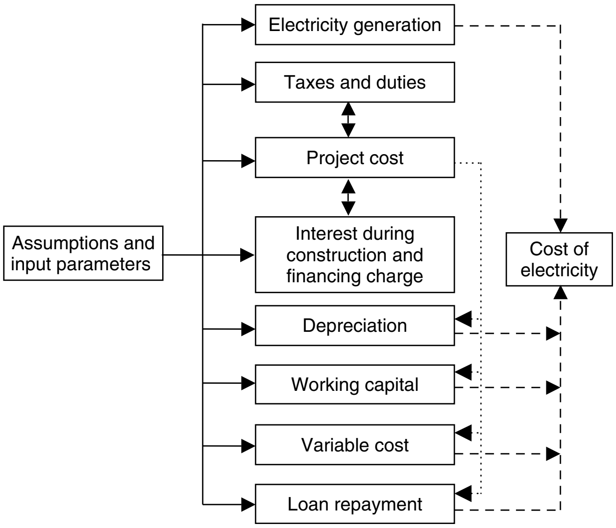


Fig. 1 Components contributing to LCC of electricity for coal and wind power projects.

$$E_{PV} = 356 \times ESH \times PV_{cap} \times \eta_{inv} \times (1 - f_{T,d}) \times (1 - f_w) \quad (1b)$$

PV_{cap} is the capacity of the PV system in kW, η_{inv} is inverter efficiency, and $f_{T,d}$ and f_w are the fraction of the rated module output adjusted for temperature dust load, and wire loss, respectively. The values of these parameters are given in Appendix B. The ESH varies with the system tilt angle. It is usually least at 90° system tilt, the maximum being at the latitude angle tilt. However, ESH available at the latitude +15° angle has been used in the study as, for this panel tilt, the solar energy available across each month in a year is (generally) uniform.[2]

Depreciation

Depreciation of a physical asset relates to the loss in its value over time. For energy projects, depreciation is usually assessed based on the accounting concept that the depreciated value of the asset is allocated over years and treated as an operating expense. Thus, the annual depreciated cost of an asset is used for allocation rather than valuation.[9] The method for depreciation calculation could be either: (i) straight-line depreciation; (ii) declining balance depreciation; (iii) sum-of-the-year digit depreciation; or (iv) modified accelerated capital recovery system (MACRS). In this study, the MACRS method has been used to calculate asset depreciation. The depreciation rate is likely to vary for different assets of the plant. For example, electrical components usually attract higher depreciation rates as compared to mechanical systems. The annual depreciation of the assets, DEP, can be expressed as:

$$DEP = \sum (PC_{P\&M} + PC_{CW} + PC_{MFA}) \times r_{dep} \leq PC \times f_{dep} \quad (2)$$

$PC_{P\&M}$, PC_{CW}, and PC_{MFA} (in \$) represent the asset components of the project cost (PC) for plant and machinery, civil works, and miscellaneous fixed assets, respectively. The depreciation rate is represented by r_{dep}. As mentioned above, different depreciation rates are likely to be applicable for the three categories of the depreciable assets. However, in this study, all assets are depreciated with the same annual depreciation rate. f_{dep} is the depreciable fraction of the project cost.

Financing Charge

During the project construction and commissioning period, a part of the project cost is funded through equity and the balance through a loan. The amount of the loan drawdown depends on the equipment and services that have been supplied during the construction of the project. And the corresponding interest on the loan is compounded on a monthly basis. Financing charges are usually associated with the loan amounts and categorized as[4,10]:

1. Management agreement fee
2. Commitment fee
3. Bank guarantee commission
4. Bank guarantee fee
5. Financial advisory fee

The financing charges, as a fraction of the loan, may be a one-time upfront charge or a recurring charge based on periodic drawdown of loan amounts. For example, the management agreement fee, bank guarantee fee, financial advisory fee are one-time upfront charges.[10] Charges accrued on bank guarantee commission and commitment fee are progressive in nature. Mathematically, the financing charge (FC) can be expressed as:

$$FC = \sum_k FC_k = \left[\underbrace{(CC_{i,j} + NCC_{i,g}) \times f_l \times f_m}_{\text{Upfront finance charge}} + \underbrace{\left[\sum_{k=1}^{n} (CC_{i,j} \times f_{i,j,k,n}) + \sum_{k=1}^{n} (NCC_{i,g} \times f_{i,g,k,n}) \right] \times f_l \times f_o}_{\text{Progressive finance charge}} \right] \quad (3)$$

where $CC_{i,j}$ is the capital cost corresponding to ith project for its jth component—for example, the wind turbine (the jth component) for the wind power project (the ith project). $NCC_{i,g}$ represents the noncapital cost of the ith project for

Inv–Light

its gth component (e.g., legal expenses). f_l is the fraction of $CC_{i,j}$ and $NCC_{i,g}$ funded through the loan. A fraction of f_l is f_m, representing the mth upfront financing charge. $f_{i,j,k,n}$ is the fraction of $C_{i,j}$ supplied in the kth month and n is the maximum number of months required for its construction and commissioning. $f_{i,g,k,n}$ is the fraction of $NCC_{i,g}$ phased in the kth month and f_o represents a fraction of f_l associated with the oth progressive financing charge.

Equity finance is the provision of money or goods and services that gives partial ownership of assets and liabilities of a project to an investor(s). Large corporations normally raise equity by selling shares and bonds in the capital market and also through allotment of preferential shares to investors. The structure of equity ownership, its concentration, and composition are crucial for equity investors. The composition of the equity investor can be an individual, a family or family group, a holding company, a bank or an institutional investor like an investment company, a pension fund, an insurance company, or a mutual fund.[1] Equity investors require a minimum return on investment for their capital and they claim residual surpluses from project cash inflows.

Debt is the amount of money loaned by a party to another party at a mutually agreed repayment terms. Debt is characterized by several attributes, such as repayment period (or maturity), repayment provisions, seniority, security, and interest rates. Debt maturity can be characterized in terms of short-term (up to 1 year), medium-term (up to 5 years), and long-term (more than 5–7 years) loan repayment periods.

Interest During Construction

Interest during construction is the accumulated monthly interest on the loan used to finance construction and commissioning activities of a project. For example, a large coal power project takes around three or more years for its construction and to become operational. The IDC will depend on the phasing of equipment and services required during construction of the project and the corresponding loan drawdown to finance the same. Therefore, the IDC can be expressed as:

$$\mathrm{IDC} = \sum \left\{ \left[\sum_{k=1}^{n} (CC_{i,j} \times f_{i,j,k,n}) + \sum_{k=1}^{n} (\mathrm{NCC}_{i,g} \times f_{i,g,k,n}) \right] \times f_l + \mathrm{FC}_k \right\} \times \left(\frac{I_J}{12} \right) \quad (4)$$

where I_J is the annual interest rate on J sources of the loan. It is normal for large power projects to secure loans from more than one source.

Project Cost

The total cost of a power project considers three important components and these are: (i) the project's capital and noncapital costs; (ii) the IDC; and (iii) the financing cost. The project cost can be calculated as:

$$\mathrm{PC} = \left(\sum \mathrm{CC}_i + \sum \mathrm{NCC}_i \right) + \mathrm{IDC} + \mathrm{FC} \quad (5)$$

Variable Cost

The variable costs associated with a plant are typically the cost of consumables and the "operation and maintenance" costs. For a coal power plant, the consumables are coal, oil, and water, whereas these costs' components are not applicable in the case of wind and solar power projects. The variable cost can be expressed as:

$$\mathrm{VC} = \sum [\{(S_{\mathrm{coal}} \times C_{\mathrm{coal}}) + (S_{\mathrm{oil}} \times \rho_{\mathrm{oil}} \times C_{\mathrm{oil}})\} \times E_i + (8760 \times \mathrm{plf} \times S_{\mathrm{water}} \times C_{\mathrm{water}}) + (\mathrm{CC}_i \times f_{\mathrm{O\&M}} \times f_{\mathrm{vc}})] \times (1 + e)^t \quad (6)$$

where S_{coal}, S_{oil}, and S_{water} are specific coal, oil, and water consumptions, respectively. The cost of coal, secondary oil, and water are represented as C_{coal} (in \$/t), C_{oil} (in \$/t), and C_{water} (in \$/1000 m^3/h). The density of oil is ρ_{oil}. It is worth mentioning that even in a coal power plant, oil (or, as it's usually termed, secondary oil) is used as a fuel during the boiler startup. In calculating the variable cost, a fraction of the capital cost is assumed to meet the operation and maintenance of the plant ($f_{\mathrm{O\&M}}$). Further, this fraction is subdivided into components contributing to the variable cost (f_{vc}) and fixed cost. The annual average plant load factor for the coal power plant is given as plf. The annual inflation rate is represented as e in the tth year.

Interest on Working Capital

In order to maintain uninterrupted operation of a large power plant, some working capital is required to meet the costs arising out of unanticipated circumstances. For example, in a coal power plant, a few months of consumables (coal and oil) would always be stocked in case of temporary disruption in fuel supply. In addition, costs associated with debt servicing (or the loan repayment) and operation and maintenance for a few months are also considered in the working capital. The working capital is generally funded through very short-term loans (i.e., loans for a few months). The interest rate applicable to the working capital is likely to be higher than

Inv–Light

that for long-term loans. The total interest on working capital (I_{WC}) can be calculated as:

$$I_{WC} = i_{wc} \times \left[\sum \left(VC_i \times \frac{\beta_i}{12}\right) + LR \times \frac{\gamma}{12}\right] \quad (7)$$

where i_{wc} is the annual interest on working capital and VC_i is the ith component of the variable cost (e.g., costs of coal and oil and the "operation and maintenance" cost) for which the provision of working capital is required. β_i is the number of months of working capital needed, corresponding to VC_i. Loan repayment is the value of annual loan repayment and γ the number of months for which the working capital for LR is required.

Taxes and Duties

The taxes and duties are applicable to different components of the project's capital and noncapital costs. Import duty is levied on part of the goods and services procured from overseas. Local, state, and federal taxes and duties, such as sales tax, apply to the equipment and services required for the project. Mathematically, the costs of taxes and duties can be expressed as:

$$TD = \sum (CC_i + NCC_i) \times R_{TD,i} \quad (8)$$

where $R_{TD,i}$ is the rate of tax or duty for the ith component of capital or noncapital costs.

Loan Repayment

As mentioned before, a certain fraction (f_d) of the total project cost is funded by debt (D) and the rest by equity. However, it is likely that large projects raise debt from more than one source (J sources) such that:

$$D = f_d \times PC = \sum D_J \quad (9a)$$

$$LR = \sum LR_J = \sum_q \left\{ \underbrace{\left[\frac{D_J}{T_{L,J} \times q_J}\right]}_{\text{Principal amount}} + \underbrace{\frac{D_J \times I_J}{q_J}}_{\text{Interest amount}} \right\} \quad (9b)$$

LR is the annual loan repayment amount. $T_{L,J}$ is the period of the loan repayment (in years) and I_J is the interest on Jth loan. q_J is the number of loan repayments in a year. In Eq. 9b, the principal payment remains constant throughout the term of the loan period. However, the interest is paid on the declining balance of the loan amount (D_J). Therefore, in the loan repayment proceeds, the interest amount is high during the initial years of the loan term. Toward the final years of the loan term, the principal amount is the major component of the loan repayment.

Life-Cycle Cost of Electricity

The LCC of the electricity (in \$/kWh) of a project is obtained by taking all of the associated costs within the life of the project (N) into account and adjusting with a discount rate (d). The selection of a suitable discount rate is important for LCC analysis since a high discount rate will tend to favor options with low capital costs, short life spans, and high recurring costs, while a low discount rate will have the opposite effect (Ref. 13, p. 338). The LCC of electricity can be expressed as:

$$E_{LCC} = \sum_{t=1}^{N} \left\{ \frac{1}{E_i} \frac{1}{[1 + (d - e)]^{t-1}} \times [ROE + DEP + VC + I_{WC} + LR + TD] \right\} \quad (10a)$$

$$E_{LCC,PV} = \sum_{t=1}^{N} \frac{1}{E_{PV}} \frac{1}{[1 + (d - e)]^{(t-1)}} \times (C_{sys,PV} + C_{Rpl} + f_{O\&M-PV} C_{sys,PV}) \quad (10b)$$

where E_{LCC} is the LCC of electricity from the wind or the coal based power projects. $E_{LCC,PV}$ is the LCC of electricity from the PV system. ROE is the return on equity invested into the project. An investor is likely to expect minimum return on investment (as equity) into the power project. In the present study, the equity return is assumed at 16%. $C_{sys,PV}$ and C_{Rpl}, respectively, are the capital cost of the PV system and the replacement cost of equipment within the system life (i.e., inverter replacement every 10 years). $f_{O\&M-PV}$ is the fraction of the system's capital cost assumed to meet the annual operation and maintenance costs associated with the system (see Appendix B).

Financial and Fiscal Incentives Offered on Renewable Energy Systems

Eq. 10 shows all of the costs associated with the energy project/plant added and discounted (with an inflation adjusted real discount rate) to obtain the present value of the LCC. However, any financial or fiscal incentive offered to an energy system also needs to be adjusted in its LCC. In recent years, several renewable energy (RE) technologies were aided with financial and fiscal incentives (FIs) offered by the federal and respective state governments in order to mainstream RE as a viable option (see Ref. 5 for an in-depth analysis of the effects of FIs on the users of solar energy technologies). In the United States, about 22 states have adopted the "Renewable Energy Portfolio Standard" (RPS) or a similar framework, (see www.dsireusa.org for an overview of federal and

Table 3 Financial incentive on PV and wind energy systems

Technology	Sector	Federal incentives and the incentive level	State incentives and the incentive level
PV	Residential	Income tax credit for maximum of \$2000 or 30% of the system cost	Property tax exemption: for 20 years assumed at 1.5% of the system cost per year Capital subsidy: offered by the MTC at \$2/kW$_{dc}$ Capital subsidy: offered by MTC for state (Massachusetts) manufactured components at \$0.5/kW$_{dc}$ 100% sales tax exemption Income tax credit at maximum of \$1000 Production credit: sale of renewable energy certificates at 6 cents/kWh offered by mass Energy
Wind	Commercial / Industrial	Modified accelerated cost recovery system—100% of the total asset depreciation in five years, using MACRS rates Production tax credit: renewable energy certificate (REC) at 1.9 cents/kWh for first 10 years from 2005 (with inflation adjusted in the subsequent years). Also, the REC sale is assumed for rest of the plant life	Property tax exemption: for 20 years assumed at 1.5% of the project cost per year 100% sales tax exemption

The values of the incentives have been used from www.dsireusa.org (accessed on 09/05/2006). However, note that the incentive levels change from time to time.

state financial incentives on RE technologies) which has been emerging as an effective mechanism to increase the share of RE in a state's energy mix. Typically, the financial incentives offered by the federal agencies on RE technologies are: the accelerated depreciation for cost recovery, renewable electricity production tax credit, etc. The state-offered FIs are: sales tax rebate, income tax credit, property tax exemption, system cost buy-down rebate, etc. In addition, the applicability of FIs vary considerably among the RE technologies, as well as among the sectors (i.e. residential, commercial, and industrial sectors) and the states in which the RE systems (or projects) are implemented.[2] Therefore, to illustrate the effects of FIs on LCC, we consider the incentives applicable in the state of Massachusetts on PV and wind-based RE technologies. In addition, the federal incentives applicable to the PV and wind-based energy systems have also been considered (Table 3). The present value of the various FIs can be calculated as:

$$ST = CC_i \times r_{ST} \tag{11a}$$

where ST (in \$) is the present value of the state sales tax exemption on purchase of RE equipment and systems. This is assumed to apply on the capital cost of the respective energy system or project (i.e., PV system and wind power project). r_{ST} is the sales tax rate applicable in a state. For MA, its value is 5% of the capital cost.

$$PT = \sum_{t=1}^{N_{PT}} \frac{1}{[1 + (d - e)]^{t-1}} \times PC \times \left[1 - \frac{t}{N_{PT}}\right] \times r_{PT} \tag{11b}$$

where PT (in \$) is the present value of annual property tax assessed on the installed RE project/system. Note that the property tax is assessed annually on a diminishing value of the project cost (PC) by a factor $[1-t/N_{PT}]$. N_{PT} is the maximum number of years the project is exempted from the property tax payment (which is 20 years in Massachusetts, as shown in Table 3) and r_{PT} is the property tax rate.

$$CS_{PV} = PV_{cap} \times (r_{CS-sys} + r_{CS-comp}) \tag{11c}$$

CS_{PV} (in \$) is the capital subsidy (or, as commonly referred to, the system buy-down rebate) offered on a PV system. Capital subsidy is usually offered as an upfront rebate on the rated capacity of the PV system. r_{CS-sys} and $r_{CS-comp}$ (in \$/kW$_{DC}$ or \$/kW$_{AC}$) are the subsidy rates on system and components manufactured respectively in the state (Table 3).

$$IR = (IR_{state} + IR_{fed}) \tag{11d}$$

where IR (in \$) is the income tax rebate offered to a buyer of a PV system. The IR_{state} and IR_{fed} are the dollar amounts of the income tax rebate offered by state and federal agencies, respectively (Table 3). In this study, the income tax rebate is assumed to be availed by the buyer of the system in the very first year it is purchased. However, the income tax rebate is generally offered with the provision to carry it forward for more than one year (usually for 3–5 years).

$$\text{DEP}_{\text{ACCL}} = \sum_{t=1}^{N_{\text{ACCL-DEP}}} \frac{1}{[1 + (d - e)]^{t-1}} \times \text{PC} \times r_{\text{ACCL-DEP},t} \tag{11e}$$

Accelerated depreciation of the project assets has been a useful mechanism that allows a commercial (and/or an industrial) entity an early recovery of the project costs. For example, the current provisions allow 100% project cost recovery within 5 years of project operation by applying accelerated depreciation rates. In Eq. 11e, DEP_{ACCL} (\$) is the present worth of the depreciated value of the project cost within a period of $N_{\text{ACCL-DEP}}$ (years) with a depreciation rate of $r_{\text{ACCL-DEP},t}$, applicable in the tth year.

$$\begin{aligned} \text{REC}_{\text{Wind}} = & \sum_{t_1=1}^{N_{\text{REC}}} \left\{ \frac{1}{[1 + (d - e)]^{t_1-1}} \times \left(E_i \times \frac{P_{\text{REC},1}}{100} \right) \times (1 + e)^{t_1} \right\} \\ & + \sum_{t_2=N_{\text{REC}}+1}^{N} \left\{ \frac{1}{[1 + (d - e)]^{t_2-1}} \times \left[E_i \times \frac{P_{\text{REC},2}}{100} \right] \times (1 + e)^{t_2} \right\} \end{aligned} \tag{11f}$$

where REC (\$) is the present value of renewable energy certificates (RECs) generated from the wind power project. The capped price of an REC unit is 1.9 cents/kWh, starting from 2005, and it is adjusted with the annual inflation rate in the subsequent years. $P_{\text{REC},1}$ is the unit REC price (cents/kWh) for period (t_1) of the first 10 years of the plant's operation (Table 3) and is a tax exempt revenue source. It is assumed that the RECs generated beyond the first 10 years to the end of the plant life (period t_2) could also be sold at the unit price of $P_{\text{REC},2}$; however, revenue generated from REC sale is taxable. In this study, the unit price for REC for both periods is assumed to be the same. The RECs generated from the PV system can also be sold, however, at a unit rate different from that of a wind power plant and can be expressed as:

$$\text{REC}_{\text{PV}} = \sum_{t=1}^{N} \left\{ \frac{1}{[1 + (d - e)]^{t-1}} \times \left(E_s \times \frac{P_{\text{REC,PV}}}{100} \right) \times (1 + e)^{t} \right\} \tag{11g}$$

where $P_{\text{REC,PV}}$ (in cents/kWh) is the price of a REC unit generated from the PV system.

The financial advantage of installing a PV system to generate electricity on a residential site is that it partially offsets the cost of electricity purchased from grid. Therefore, the energy costs saved due to the PV system can be obtained as:

$$\text{ES} = \sum_{t=1}^{N} \frac{1}{[1 + (d - e)]^{t-1}} \times E_{\text{PV}} \times T_{\text{R}} \tag{11h}$$

where ES is the cost of the grid electricity avoided due to on-site electricity generation and T_{R} is the electricity tariff applicable in the residential sector. Therefore, the effect of financial incentives on the life cycle of electricity from wind and PV power systems can be expressed as:

$$E_{\text{LCC,Incetives}} = E_{\text{LCC}} - \left(\text{ST} + \text{PT} + \frac{\text{REC}_{\text{Wind}}}{E_i} + (\text{DEP}_{\text{ACCL}} - \text{DEP}) \right) \tag{11i}$$

$$E_{\text{LCC-PV,Incetives}} = E_{\text{LCC-PV}} - \left(\text{ST} + \text{PT} + \frac{\text{REC}_{\text{PV}}}{E_{\text{PV}}} + \text{IR} + \text{CS}_{\text{PV}} + \text{ES} \right) \tag{11j}$$

RESULTS AND DISCUSSIONS

The cost profile of coal, wind, and PV power systems are different since each system operates in a different manner. For instance, the installed cost of coal and wind-based projects are comparable while that for a PV system is high. As shown in Table 4, the installed cost of the coal and wind power project are \$1210/kW and \$1353/kW. For the PV system, it is \$7840/kW. The costs, input parameters, and associated assumptions are given in Appendices A–D.

It can also be noticed from Table 3 that for the coal power project, the IDC and financing charge comprise around 16% of the total project, while the corresponding value for the wind power project is less than 10%. This is due to the fact that coal power projects take a longer time to implement than wind power projects. This also indicates the importance of commissioning a project on time and within the budget. To reduce the IDC, a project developer would ideally prefer to invest the equity funds before the loan. This provides banks additional security for loans offered to project developers. In fact, for large green-field power projects, banks normally allow loan drawdown after a substantial part of the project equity funds have been invested. This is because after investing a large part of the equity, a project developer's commitment to complete the project on time and budget is likely to increase.

On the other hand, a project developer would prefer to schedule payments to the project's construction and commissioning contractors as close to the project completion schedule as possible to reduce the IDC. However, this would increase financial risks for the contractors. Usually, the project contracts are designed such that a

Table 4 Cost components of coal, wind, and PV power systems

Cost Category	Coal (million $)	Wind (million $)	PV ($)
Capital cost	445.5	49.9	23,520
Non-capital cost	86.0	11.8	—
Interest during construction	61.5	2.1	—
Financing charge	36.2	3.8	—
Total project cost	629.2	67.7	23,520
Project capacity	520 MW	50 MW	3 kW
Unit cost	1210 $/kW	1353 $/kW	7840 $/kW

certain percentage (10%–15%) of the contractual amount is paid to the contractor as mobilization advances to commence activities. The balance payments are made progressively, based on achieving predetermined project construction and commissioning milestones.

The annual electricity tariff for coal and wind power plants follow a similar trajectory, except that the electricity tariff for a wind power plant is higher than that for the coal power plant, as shown in Fig. 2a and b.

For both coal and wind power projects, the sharp decline of the electricity tariff in the 11th year is due to the full amortization of loan repayments. In the 15th year, the deprecation costs are also fully adjusted. Thereafter, the funds required to meet the fixed charges reduce substantially. From the 15th year onward, the components of the variable charge contribute largely to the electricity tariff. Because wind power projects do not require fuel during operation, the electricity tariff is predominantly dictated by the fixed cost components of the project. On the other hand, for coal power plants, the components of variable charge contribute largely after the loan repayment and depreciation costs are fully adjusted.

The levelized cost or the LCC of electricity is shown in Fig. 3 for the three power projects considered in this study. The costs, input parameters, and assumptions are given in Appendices A–D. As expected, the LCC of electricity from a coal power plant is the least expensive, followed by wind, while it is highest for the PV system. It is interesting to note that the installed cost (in $/kW) of a solar system is about six times higher than that for the coal power project (Table 3). However, the levelized cost of electricity from these two sources differs by around ten times. This effect is due to the different capacity factors utilized by these systems. A coal power plant usually operates around 80% of the time in a year, while a PV system operates only

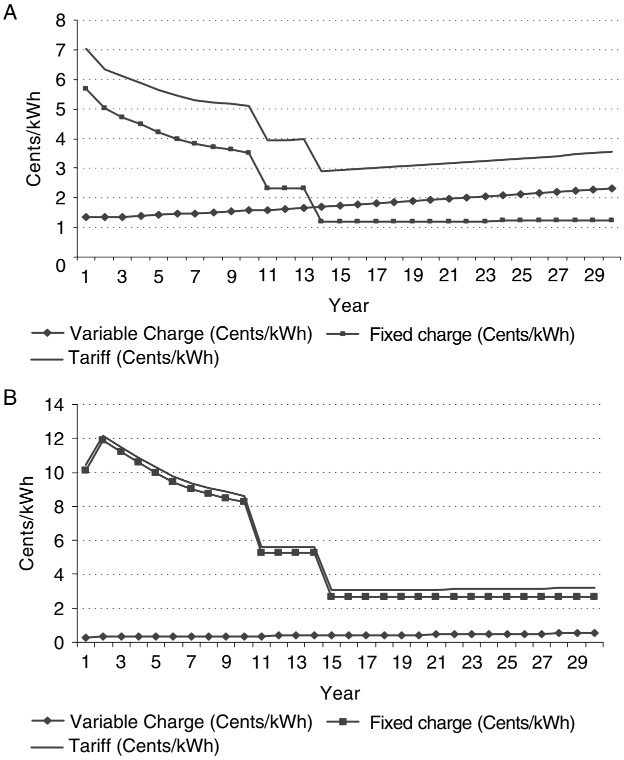


Fig. 2 (A) Annual electricity cost for the 520-MW coal plant. (B) Annual electricity cost for the 50-MW wind plant.

around 15% in a year. The levelized cost of electricity from the wind power plant is around twice that of the coal power project and the corresponding values are 4.4 and 7.6 cents/kWh, respectively.

From the above discussion, it can be inferred that the cost of electricity associated with wind and PV-based power systems can possibly impede investors or individuals from installing such systems for electricity generation. In order to promote the use of electricity from RE sources, several state and federal agencies in the United States (and several other countries) have been offering FIs for RE technologies. Typically, the state-offered FIs consist of sales tax exemption on the purchase of system components, property tax exemption, income tax rebate, capital subsidy on the system cost, etc. The FIs offered by federal agencies are typically the accelerated depreciation on the project cost for faster recovery of the capital, the production tax incentive, income tax rebate, etc.

To illustrate the effects of FIs on the LCC of electricity from the PV system and the wind power project considered in this study, the federal and Massachusetts state financial (and fiscal) incentives offered on residential PV system and utility scale wind power projects have been used.

The FIs have a significant effect on the LCC of electricity generated from the PV system. The capital subsidy on the system, along with the electricity costs saved due to onsite power generation, contributes to a

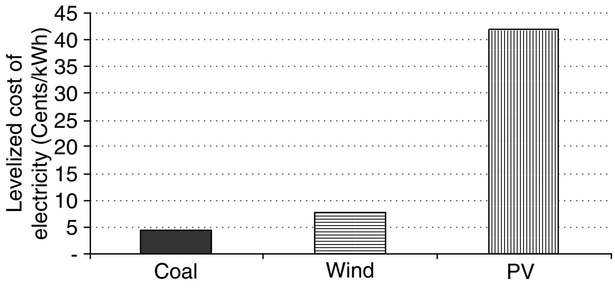


Fig. 3 Levelized cost of electricity from coal, wind, and PV-based power systems.

reduction of as much as 65% of the LCC of electricity from the system (Fig. 4). When several other FIs, such as federal and state income tax rebates, state sales and property tax exemptions, and the sale of RE credits generated from the electricity produced by the system are considered, the LCC of electricity from the PV system is negative (Fig. 4).

Similarly, the LCC of electricity from the 50 MW wind power project is significantly reduced when the effect of FIs is considered. The accelerated depreciation for capital cost recovery and the production tax credit contribute to reducing the cost of electricity by 65% (Fig. 5). The LCC of electricity from the wind power project without the FIs is 7.6 cents/kWh and when the effect of FIs is considered, the corresponding value is 2.1 cents/kWh (Fig. 5).

The cost of electricity from coal, wind, and PV-based power systems after considering the effects of FIs are compared in Fig. 6. The cost of electricity generated from the PV-and wind-based energy systems could be significantly less than that generated from coal power projects (contrary to the conventional belief; however, the financial incentives on PV and wind-based RE systems vary considerably from state to state. Therefore, the results of this study are only illustrative in nature). This is (possibly) one of the reason for large-scale deployment of residential PV systems and utility scale wind power projects in several parts of the United States (and elsewhere) in the recent years.

CONCLUSIONS

Life-cycle cost assessment provides an estimate of the major expected costs that can be incurred within the useful life of an asset. Life-cycle cost estimation is a tool that influences investment decisions for acquiring or developing an asset. It allows for the comparison of different investment alternatives and thus enables determination of the most cost-effective system. This study has attempted to develop a framework for conducting LCC analysis of energy projects by focusing on two phases of a project's life-cycle costs, these are

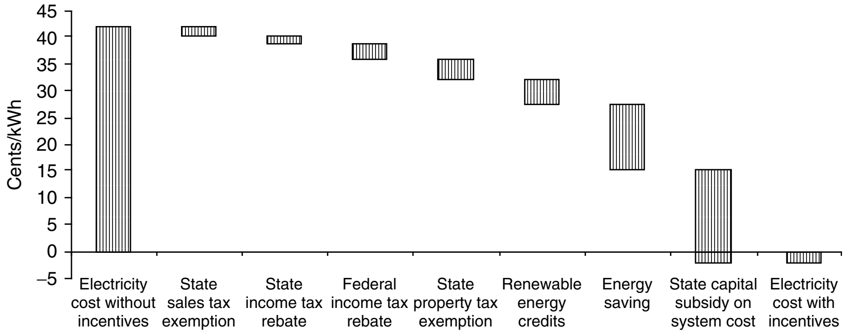


Fig. 4 Effect of financial incentives on the levelized cost of electricity from the 3-kW PV system.

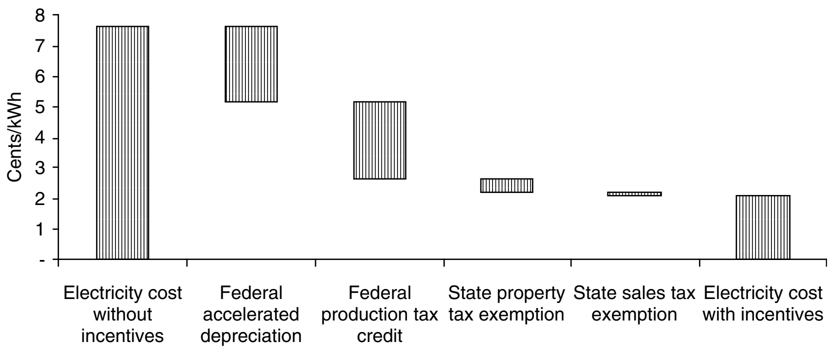


Fig. 5 Effect of financial incentives on the levelized cost of electricity from the 50-MW wind power project.

Inv–Light

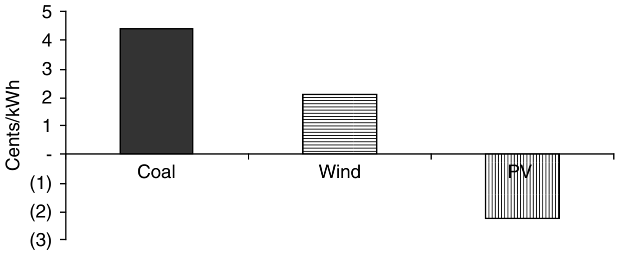


Fig. 6 Levelized cost of electricity with financial incentives from coal, wind, and PV-based power systems.

(i) the development, construction, and commissioning phase and (ii) the operation phase. The energy projects considered were: (i) a 520-MW coal power plant; (ii) a 50-MW wind power project; and (iii) a 3-kWp grid-tied residential PV system. The LCC assessment considers costs associated with the debt servicing, variable costs, interest on working capital, depreciation, annual operation and maintenance cost, etc. Most of these costs are applicable for large (or utility scale) power projects and the effect of these costs on the electricity produced from a grid-tied residential PV system is negligible. The LCC of electricity from the PV system is around ten times higher than that from the coal power source when the effects of financial incentives are not considered. Similarly, the LCC of electricity from the wind power project is around twice that from the coal power plant without considering the FIs. The annual cost of electricity generated from a coal power plant is balanced between fixed and variable costs, while that for the wind power project is largely contributed by fixed cost.

In addition, a framework has been developed to quantify the financial and fiscal incentives offered by federal and state agencies to RE technologies. This enables studying the effect of FIs on the LCC of electricity generated from the PV and wind power systems. For the residential PV system, the state offered upfront capital subsidy on the system cost and the cost of electricity that was avoided due to the contribution of onsite power generation in reducing electricity cost (LCC) by 65%. Similarly, the LCC of electricity from the 50-MW wind power project is significantly reduced when the effect of FIs is considered. The accelerated depreciation for capital cost recovery and the production tax credit contribute to reducing the cost of electricity by 65%. The cost of electricity generated from the PV and wind-based energy systems could be significantly less than that generated from coal power projects when the effects of financial incentives are considered in the calculation of the LCC of electric power projects.

ACKNOWLEDGMENT

I am grateful to Gargi Bagchi for critical comments on the manuscript.

APPENDIX A: ASSUMPTIONS AND INPUT PARAMETER

Particulars	Unit	Coal	Wind
Gross capacity	MW	520	50
Auxiliary consumption	%	9.5	—
Plant capacity factor	%	80	30%
Specific oil consumption	ml/kWh	3.5	—
Density of secondary oil	gm/cc	0.85	—
Water consumption	cum/h	3000	—
Construction period	Months	42	12
Debt	%	70	70
Equity	%	30	30
Depreciation for 90% of project cost using the MACRS	%	90	90
Working capital requirements			
Coal cost	Months	2.5	
Secondary oil stock	Months	2	
Coal stock	Months	0.5	
O&M expenses	Months	2.5	2.5
Debt service	Months	1	1
Annual inflation rate	%	5	5
Coal cost	$/t	12.8	
Taxes and duties (sales tax)	%	5	5
Corporate tax rate	%	35	35
Interest on working capital	%	12	12
Useful plant life	Years	30	30
Loan repayment period	Years	10	10
Cost of secondary oil	$/t	162.8	—

(Continued)

Inv–Light

(*Continued*)

O&M cost, including insurance charges (as % of capital cost)	%	2.5	2.5
Financing charges			
Management agreement fee	%	1.125	1.125
Commitment fee	%	0.075	0.075
Bank guarantee commission	%	1.6	1.6
Bank guarantee fee	%	3.15	3.15
Financial advisory fee	%	1.5	1.5

Source: From Elsevier Science Ltd. (see Ref. 3).

APPENDIX B: INPUT VALUES FOR LIFE-CYCLE COST ANALYSIS OF A 3-KWP PV SYSTEM

Parameters	Unit	Value
Wire losses	Fraction	0.02
Inverter cost	$/kW	885
PV system life	Year	25
PV system size	kWp	3
PV system cost	$/kW	7840
Inverter efficiency	%	95
Inverter replacement	Year	10
Annual inflation rate	%	2
Annual land lease charges	$	0
Residential electricity tariff	$/kWh	0.121
Loan as fraction of the system cost	Fraction	0
O&M expenses as fraction of the capital cost	Fraction	0.02
Annual insurance cost as fraction of capital cost	Fraction	0.005
Property tax as fraction of residual value of system cost	Fraction	0.015
Temperature and dust loses as fraction of module rated output	Fraction	0.1
PV system deprecation for book value (by the MACRS method)	Year	15

Source: From International Solar Energy Society (see Ref. 2).

APPENDIX C: CAPITAL AND NONCAPITAL COSTS OF A 50-MW WIND PROJECT

	Wind (in million $)
Capital costs	
Turbine and tower	33.23
Foundations	0.35
Assembly and checkout	0.19
Electrical infrastructure	1.47
Substation	0.25
Roads and grading	0.4
Control systems	0.4
Control buildings	0.1
Central building	0.1
Transmission upgrade	2.2
Noncapital cost	
Initial expenses	2.8
Engineering and overhead	3.2
Duties and taxes	2.2

Source: From Department of Business, Economic Development and Tourism (see Ref. 7).

APPENDIX D: CAPITAL AND NONCAPITAL COSTS OF A 520-MW COAL POWER PROJECT

	Coal (in million $)
Capital Cost	
Steam generator island	103.3
Turbine generator island	64.5
Mechanical systems	
Coal handling plant	25.3
Ash handling plant	20.0
Cooling tower and circulating water system	7.3
Miscellaneous systems	42.5
Electrical systems	—
Generator transformer	4.0
Switchyard	16.1
Miscellaneous systems	14.0
Control and instrumentation	10.1
Intake channel/pipes and related equipment	7.8
External coal transportation system	17.9
Initial spares	10.3

(*Continued*)

Inv–Light

(*Continued*)

Civil works	
Access and diversion road	1.3
Ash disposal area development	10.5
Staff residential facility	4.0
Temporary construction	11.6
In-plant civil works	45.9
Miscellaneous fixed assets	1.7
Noncapital cost	
Erection, testing, and commissioning	40.1
Construction insurance	4.1
Engineering and overheads	23.4
Preliminary expenses	2.8
Land and site development	4.2
Owner's engineers expenses	2.0
Operator's training	0.7
Start-up fuel	1.8
Legal expenses	2.1
Establishment cost	4.1
Duties and taxes	17.2

Source: From Elsevier Science Ltd. (see Ref. 3).

REFERENCES

1. ADB. An Analytical Framework of Corporate Governance and Finance. In *Corporate governance and finance in East Asia: A study of Indonesia, Republic of Korea, Malaysia, and Thailand*, Vol. 2; Asian Development Bank: Manila, 2001; [Chapter II].
2. Bhattacharjee, U.; Duffy, J. Renewable energy portfolio standard: impacts of state financial incentives on residential owners of a photovoltaic system. *International Solar Energy Society*, Proceedings of the Solar World Congress, Denver, CO, U.S.A., 2005.
3. Bhattacharjee, U.; Kandpal, T.C. Potential of fly ash utilization in India. Energy **2002**, *27* (2).
4. Bhattacharjee, U.; Kandpal, T.C. Financial evaluation of coal thermal power generation in India: effect of incremental cost of fly ash utilization. Proceedings of the World Renewable Energy Congress VII, Cologne, Germany, 2002.
5. Chandrasekar, B.; Kandpal, T.C. Effect of financial and fiscal incentives on the effective capital cost of solar energy technologies to the user. Solar Energy **2005**, *78* (2), 147–156.
6. Fthenakis, V.M. End-of-life management and recycling of PV modules. Energy Policy **2000**, *28*, 1051–1058.
7. Hawaii (2004). Select Hawaii Renewable Energy Project Cost and Performance Estimates, 2004, The State of Hawaii, Department of Business, Economic Development & Tourism. Strategic Industries Division. Source: http://www.hawaii.gov/dbedt/info/energy/publications/
8. IEA. *Energy Technologies at the Cutting Edge*; International Energy Agency: Paris, 2005.
9. Kandpal, T.C.; Garg, H.P. *Financial Evaluation of Renewable Energy Technologies*; Macmillan India Limited: Delhi, 2003.
10. Khalaf, A.G.; Redha, M.A. A case study in formulating financial modeling for evaluation of levelized unit cost of a new power and water plant for the ministry of electricity and water, Kingdom of Bahrain. Desalination **2002**, *152*, 67–74.
11. NYSERDA (2005a). Wind Energy Project Lifecycle Overview. Wind Power Toolkit. Source: http://www.powernaturally.org/Programs/Wind/toolkit.asp.
12. NYSERDA (2005b). Economic and Socioeconomic Impacts of Utility-Scale Wind Power Projects. Wind Power Toolkit. Source: http://www.powernaturally.org/Programs/Wind/toolkit.asp.
13. Woodward, D.G. Life cycle costing—theory, information acquisition and application. Int. J. Project Management **1997**, *15* (6), 335–344.

BIBLIOGRAPHY

1. Gluch, P.; Baumann, H. The life cycle costing (LCC) approach: a conceptual discussion of its usefulness for environmental decision-making. Building and Environment **2004**, *39*, 571–580.

Life Cycle Costing: Energy Projects

Sandra B. McCardell
Current-C Energy Systems, Inc., Mills, Wyoming, U.S.A.

Abstract
Life cycle costing for energy projects is described as a decision-making tool for energy projects, with several examples and a spreadsheet that can be used as a template for calculations. The focus of the chapter is on the principles of life cycle costing to provide managers and professionals a methodology to compare a variety of complicated projects having different costs and benefits.

INTRODUCTION

Life cycle cost analysis (LCCA) is a useful technique for comparing the relative economic benefits of several energy efficiency opportunities among which one must choose. It can also be used to assess the viability of particular investment opportunities.

The technique is one borrowed and adapted from financial analysis. As applied to energy efficiency project analysis, it has developed a particular set of procedures and approaches, and is often used for compliance enforcement as well as for its original purpose in economic decision-making.

LCCA can be simple, with few variables necessary or used, or it can be so complex that a computer model is required to examine the integrated costs and benefits of several energy efficiency systems.

This chapter is designed as a resource for those who assess energy efficiency (and other) investments. It focuses on the ways in which LCCA can be used to inform and support energy project decision-making, with an emphasis on the general approach and example spreadsheets that can be used for a simple analysis.

Computer models employ the same techniques but use the power of the computer to provide an analysis based on iterative simulations that are value-timed for periods of 20 or 30 years. A brief list of some of the more commonly used computer-modeled LCCA systems is included for those who may be interested; computer models are outside the scope of this chapter, however.

DEFINITIONS

Blended utility rate: The utility rate is the basis on which the utility charges for its products (electricity or natural gas) and its services (for example, transmission costs in the case of natural gas, power factor penalties and demand charges in the case of electricity). For general economic analysis including life cycle costing for energy projects, a "blended" utility rate should be used. A blended rate summarizes all charges for the month and then divides that number by the amount consumed. (For more detailed engineering and economic analyses, time-of-use parameters and differential quantity charges may require that the utility rate components be adjusted individually, but the blended rate gives a reasonable approximation for most purposes.)

Btu/sf: Energy use for a building is often summarized and compared on the basis of Btu/sf, or British thermal units/square foot of facility area. This useful measure converts electrical, natural gas, propane, and other energy sources to Btu and then divides the total by the number of square feet in the facility. Most "benchmark" energy comparisons use this number for total utility usage.

Energy efficiency measure (EEM): Any energy project may be composed of one or more EEMs (such as replacing single-pane windows with double-paned and appropriately glazed windows) that are individual measures designed to reduce utility costs by themselves and in concert with other EEMs.

Internal rate of return (IRR): This popular traditional financial investment comparison method evaluates projects based on all associated projected cash flows. Potential investments are normally compared either to one another or to an organization's stated target or "hurdle rate."

Life cycle analysis (LCA): LCA is a concept that is outside the scope of this chapter. It is an emerging technique designed to capture the full economic effect of the environmental costs associated with a product, project, or service, including production, transportation, end of life, and other costs.

Life cycle cost analysis (LCCA): LCCA (also known as life cycle costing) is an economic evaluation tool that is extremely useful in both evaluating and comparing energy projects. It takes into account the timing of those costs and

Keywords: Life cycle costing; Life cycle cost analysis; Economic analysis; Total owning costs; Payback period; Energy cost modeling; Building analysis.

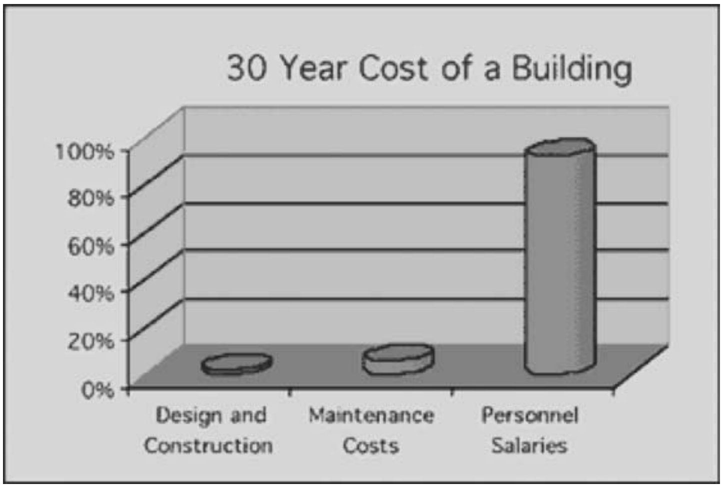


Fig. 1 Thirty-year cost of a building.
Source: From Whole Building Design Guide, Sieglinde Fuller (see Ref. 1).

benefits, as well as the associated time value of money, returning a single value for the total life cycle cost of the energy project.

Net present value (NPV): The NPV is a traditional financial analysis tool that takes all cash flows expected from a project and discounts them back to the present; positive numbers are generally seen as worthwhile investments.

Payback period: The simple payback period (PB period or simple payback) is a traditional measure used in evaluating energy projects. The calculation divides the total installation cost by the savings projected to determine the number of months it will take to pay back the investment.

Time value of money: This phrase encompasses the idea that a dollar received or spent today is worth more than a dollar received or spent in the future. Inflation is part of the reason for this; the other is that one can use the dollar today in other ways, whereas one must wait to receive and use the dollar tomorrow.

Total owning costs: This term is synonymous with life cycle cost but is used more frequently in the residential context.

THE ANTECEDENTS OF LIFE CYCLE COSTING FOR ENERGY PROJECTS

LCCA is similar to most economic decision-making tools in that it takes a variety of monetary flows and converts them to a single number, making comparisons among alternatives or an organization-designated benchmark relatively simple. As is true of all such tools, it is essentially a way to compare such divergent items as apples, oranges, and kumquats with a shared metric.

The challenge of comparing unlike items on an economic basis is one humans have faced since the beginning of time. Imagine the difficulties of the early traders who had to hold in their heads the comparative values of silks, olives, and pottery throughout their trading regions, in every season.

Money was, of course, the original method used to give similar values to disparate items, and with its introduction traders had a common way to assess the value of very different commodities (as well, of course, as providing a commonly recognized and portable medium of exchange).

As money became one of the foundations on which our civilization and commercial systems were built, other improvements were made as well: the time value of money (sometimes expressed in the phrase "A dollar today is worth more than a dollar tomorrow") was recognized; with the increasing complexity of the industrial age, project costs and benefits for a variety of complicated projects were analyzed economically; and disparate activities and items were reduced to numbers for comparison. Both of these characteristics are included in LCCA.

In the 1970s, when energy savings investments were first undertaken as stand-alone projects to reduce costs, the economic approach first used was the simple payback period (SPB period). This simple and easily understood measure simply divides the project price by the savings projected, showing that the investment would be recouped in 9 months, or 2.5, or 10 years. Although SPB periods are easy to calculate and explain, they ignore the time value of money, utility rate changes, and any operational or maintenance cost increases or decreases. In fact, they present only part of the story, and it is the author's contention that LCCA should be the preferred method. In fact for organizations interested in an economic evaluation method that also integrates itself better into their other business decisions, LCCA has come to be the norm.

Note: LCCA should be distinguished from life cycle analysis, LCA, which is the emerging trend toward accounting for each aspect of a product's life from its inception through to its disposal or recycling, including the embedded cost of the energy used to manufacture and

Table 1 Gas station example, current situation

Building type	Gas station
Date of construction	1970
Square feet	7000
Description	Concrete block construction on slab, 1–2 floors, flat roof, single-paned windows, T12 fluorescent lights with magnetic ballasts inside; HID outside; 4 leaded/unleaded vehicle pumps; 6 for truck diesel
Electricity use and yearly cost	$23,000 and 102,000 Btu/sf/yr
Natural gas use and yearly cost	$21,000 and 223,000 Btu/sf/yr
Total Btu/sf/year	325,000
Energy cost/sf/year	$7.71
Maintenance costs	$5000
Operating schedule	24 h
Major energy using systems	Heating, cooling, lights

transport the product. Although LCA draws from the legacy of LCCA, it is a different tool that focuses on the global environmental costs and benefits of a particular product.

THE FRAMEWORK OF LIFE CYCLE COSTING

The concept of life cycle costing is one that is in fact familiar to us all. When we purchase a car, for example, we do basic research on fundamentals: fuel efficiency, maintenance records, and resale value. We consider our driving patterns and the length of time we plan to have the vehicle. That is essentially an economic assessment. We overlay that information on a strategic and technical assessment. First, what do we really want the car for? Is it for work, for hauling livestock, for soft summer nights on an open highway, for transporting teams of youngsters? Then we examine how the various possibilities fit that need. What is the vehicle's performance, its reliability, its safety record? Finally (when we are being reasonable), we make our decision based on these strategic, technical, and economic grounds.

When the task is the design of an energy efficient building, the approach is a similar one. To make a reasonable decision, it is important to consider strategic, technical, and economic aspects of the purchase/construction or retrofit question. The only difference is that with the lifetime of a building being 30–50 years, these decisions are even more critical.

In fact, as shown in Fig. 1, construction costs represent only about 2% of the total cost of a building, with operations and maintenance costs (including utilities) representing 6%, and associated personnel costs representing 92%. (This last statistic also highlights the reason that energy projects usually are also designed to increase occupancy comfort; in any facilities other than heavy-industrial buildings, those occupants are the single greatest cost.)

Life cycle costing is a method not only for analyzing the reasonableness of one particular decision or group of decisions, but also for examining the advantages of one investment vis-à-vis another. To a greater or lesser extent, depending on the complications inherent in the decision, it models reality—which means that the calculation mechanism can be a scrap of paper, a simple spreadsheet, or a complicated and multifaceted computer model. The same principles apply in all cases, but each is appropriate in different circumstances.

For large projects, life cycle costing standards have been developed by ASTM International (www.astm.org) and the National Institute of Standards and Technology (NIST) (www.nist.gov), and the principles inherent in those standards are extremely useful even when the calculations are less complicated than whole-building analyses.

Stage 1: Describe the Current Situation

First, LCCA incorporates the characteristics of the current situation, to give one a picture of "what is" against which to compare. This can take many forms, as shown in Tables 1 and 2.

The level of complexity in these two analyses is very different, but the principle is the same: it is important first to know what the current situation is and then to benchmark that against similar buildings (either out of one's own experience or by using a benchmarking tool such as EnergyStar; www.energystar.gov).

In these cases, for example, it becomes evident that the gas station has an immense opportunity for energy savings, as the normal energy usage in this particular climate would be approximately 2,00,000 Btu/sf/yr. (Expressed in this way, it is a measure of utility usage across various types of energy sources.) It is also clear that although the electricity and natural gas costs are almost identical, the gas usage in Btu/sf/yr is that upon which the greatest emphasis should

Table 2 HVAC system example, current situation

System type	HVAC system
Date of installation	1980
Manufacturer and model number	Trane packaged rooftop air conditioner with natural gas heating package
Description	25 ton, 350 MBH input heater
Nameplate efficiency	Estimated EER of 6
Thermal zones	2
Utility rate schedule	Local REA, average blended electricity cost $0.87
Operating schedule	24 h
Heating & cooling season	8 months heating, 4 months cooling
Maintenance costs	$860/yr

be placed. Also, maintenance costs are quite high in relation to comparable facilities, and that information too will be useful in defining the energy projects to be recommended for the facility.

For the HVAC system, any unit with an energy efficiency rating (EER) of six is about half as efficient as it should be—which provides a great deal of scope for reduced utility costs in both heating and cooling seasons. This is particularly true because the unit runs 24 h a day, all year long, so the increased efficiency should have a significant impact.

These "current" situation descriptions (and the possible energy efficiency measures associated with them) become the first step in the LCCA—the base against which other alternatives will be compared. When the analyst thoroughly understands the current picture, the next step is to examine the facility for upgrade and efficiency possibilities.

Table 3 shows a brief example comparing the life cycle cost (or total owning cost) of an incandescent light bulb vs. a compact fluorescent (CFL) bulb. The light in each case is assumed to operate 4 h per day or 1400 h per year. The CFL lasts about 7 year, whereas the incandescent bulb lasts less than 1 year; with 9 replacements, the LCCA analysis shows the CFL, at $14.94, to be a better choice than the incandescent bulb at $44.54. (As a side note, actual electrical costs have increased, and a recent recalculation for the Denver, Colorado market produced an LCCA for the CFL of $18.76 and an LCCA for the incandescent bulb of $62.04.)

Table 3 Total owning cost of an incandescent vs a CFL light bulb

	Incandescent bulb (60 W)	15 W compact fluorescent bulb
Initial cost ($)	0.25	7.00
Annual operation ($)	7.03	1.26
7-year life cost ($)	44.29	7.94
Life cycle cost ($)	44.54	14.94

Source: From www.focusonenergy.com savings analysis worksheet (see Ref. 2).

Stage 2: Definition of the Energy Project

The next step is to define the proposed project. As with the purchase of a car, this requires the definition of strategic, technical, and economic factors. In addition, for each alternative, both immediate and long-term costs and benefits should be described and understood.

This definition is necessary for both simple stand-alone projects and for energy efficiency measures that impact strongly on multiple elements and systems in a building. In almost every case, however, there will be long-term implications for each improvement. An EEM (energy efficiency measure) of "replacing incandescent exit signs with new fixtures illuminated by LEDs," for example, does not impact the operation of the building significantly. On the other hand, there will be significant differences in maintenance and operation costs, because LED exit lights and batteries can last at least 10 years, whereas incandescent bulbs are supposed to be replaced annually.

It is precisely this type of decision that LCCA is designed to facilitate. LED exit lights are more expensive than their incandescent versions and have no effect on building operation but do significantly reduce long-term maintenance expenses. More importantly, their longer battery life makes them safer. LCCA can aggregate these costs and benefits, and provide consistent and complete information for a decision on the appropriate emergency lighting choice.

One brief note is that in most LCCA analyses, intangible costs and benefits (increased safety, lower risk, increased occupancy comfort, increased employee health, and so on) are not accounted for unless they can be quantified in some way (reduced employee sick days from better air quality and less mold, for example). It is possible and increasingly common, however, for the value of such intangible factors to be included in LCCA by weighting the variables in a statistically appropriate way, by rating the importance of several intangible factors and then grading each alternative for its performance against those factors.

When the proposed project or alternative projects have been defined (which is predominantly the responsibility of the consultants, engineers, and operations personnel), the impact of the energy efficiency improvements—both short term and long term—can be quantified. This step is the essence of LCCA, and although the details can be quite complicated, the premise is simple: Projected economic impacts for each period should be calculated, most easily by completing a table with costs and assumptions over the project life (as shown below).

In the single-technology exit light replacement example cited above, it would be important to quantify the following variables:

- Wattage difference between old lights and new LED lights
- Cost (including labor) for old lights and new LED lights
- Number of fixtures to be replaced
- Frequency of changeouts at present and cost to change lights (including labor)
- Projected frequency of changeouts and cost to change lights
- Residual value or disposal costs of both old lights and new LED lights
- Current utility rates and any projected changes over the life of the project
- Time value of money, or discount rate that should be used to convert future dollars to current ones
- Projected project cost

Even in a simple example such as the LED exit lights, other variables might usefully be included in a very large building and probably would be included in a computer model. Among them would be:

- Anticipated changes in material and/or labor costs
- Possible vandalism costs
- Inventory carrying cost for replacements

In most projects, however, these additional parameters would not provide materially different answers and, thus, are not likely to be worth the effort of research and calculation.

If exit light replacement can decrease electricity usage by approximately 70% per fixture and has long-term effects on maintenance and operations, it is easy to imagine that the picture becomes exceedingly complicated when one is looking at a whole building with integrated systems and interlocking issues.

For the architects and design engineers planning to build a new recreation center with an indoor swimming pool, for example, the technical considerations involved in defining the project are many, complicated, and interrelated, and all have financial implications. It is for cases such as this that the computer LCCA models integrating energy use and economic considerations are designed, because energy use is a key component of long-term cost; the more involved models run hourly simulations of buildingwide energy use to understand the entire economic picture.

Whether the analysis is simple or complex, however, the steps are the same. With the recreation center, the first step remains a description of the current situation (or, as here with new construction, a benchmark building). Second, the project must be defined strategically, technically, and financially.

With planned construction that is as complicated as a new recreation center, a computer model including LCCA should then be used to simulate results. For building developers who are planning to apply for U.S. Green Building Council LEED (Leadership in Energy and Environmental Design) certification, such modeling is critical, but it can be extremely useful for any facility for which construction and operating costs are constrained. With the help of computer models, the technical, energy use, and financial information provided by the model can often reduce both initial construction and long-term operating costs—although, of course, the main benefit comes from reducing the larger, long-term operating costs. For this purpose, the model should be run several times, with different assumptions or using alternative equipment, to achieve optimal whole-building results before comparing the project with other possibilities. (LCCA is a tool designed to be used as a decision-making tool for managers and operational personnel. Energy modeling, on the other hand, is best accomplished by professionals trained both in the use of the specific modeling program and in analyzing the output produced by the model.)

Stage 3: Determine and Quantify Costs and Benefits

In any analysis as potentially complicated as LCCA, it is important to include all appropriate costs and benefits. On the other hand, it is easy to go overboard in collecting information, unnecessarily increasing the complexity of the variables to be considered and, thus, delaying the decision which should be made. This is an important point because with energy projects (in contrast to most potential projects evaluated by organizations for consideration), delays do not conserve cash by reducing expenditures, but instead reduce the cash available by maintaining operating expenses at an unnecessarily high level.

The costs to be considered for an energy-project LCCA analysis generally can be described as follows:

- First costs. Depending on the project, this could be construction cost, purchase or other acquisition cost, installation costs, etc.
- Operations costs. These include the following:
  - Maintenance costs
  - Repair costs
  - Subunit replacement costs (as in the replacement of a bulb in a light fixture)
  - Utility costs
- Replacement costs
- End-of-life value or cost, which might include disposal, demolition, or residual value, depending on the project or product
- Finance charges
- Benefits and/or costs that are nonmonetary but that (if quantifiable through weighting or by some other method) are important to the decision The LCCA approach, like most other financial calculators, takes each cost or benefit into account in the year in which it occurs and then discounts or adjusts the value of those flows to the present.

For the third step, again, the appropriate variables (in addition to the design and technical data that is collected for Stage 2) will need to be defined, and they address essentially the same issues as those in the LED exit-lights example. They might include the following (using the indoor-pool example):

1. How often will the pool be occupied, and by how many people?
2. How often will there be spectators, and how many?
3. What lighting levels will be required—during the day and at night?
4. What are the normal outside weather conditions?
5. How easy is it to change lights, and what will each change cost per light?
6. What are the utility rates, including demand charges?
7. What is the projected change in utility rates over the period when the building will be occupied?
8. What is the anticipated life of the facility and its major components?
9. What is the disposal cost or residual value of short-lived building systems or components?
10. What is the projected yearly maintenance cost?

It is a very complicated picture.

When looking at a retrofit option, however—for example, a recreation center with a swimming pool already in place—the complications increases dramatically. "What is" can be known, of course, along with what is working and what is not, what is proving to operate in the way it was designed to operate, and what is causing difficulties for one reason or another. Therefore, a description of the current situation is possible, but an analysis of the issues raised by the project—even when confined to those that have energy usage implications—is quite difficult.

Naturally, most complications come from the interactions of various systems. The humidity in the pool-room air must be controlled, for example, and in a dry climate, that damp heated air might be exhausted through the locker room; heat might also be recovered from the exhaust system and used in other parts of the building. The advisability of this approach would depend on and impact the design of the air handling system, the size of the pool, the numbers of spectators, the hours of use, the pool lighting, and a number of other factors. Each important system interaction can be accounted for with a variety of the building modeling tools available on the market, and the implications for the operation of the building accounted for and understood.

LCCA allows one to take the next step, which is to assess the cost implications of the actions that might be taken to improve the situation in the future. Thus, it looks at both future costs and benefits of any energy efficiency system improvements. With whole buildings or more involved systems, the complications can mushroom because of the complexity of the building and the interactions among systems, and computer models can take such implications into account. The principle for comparison remains the same, and for most projects, a simple spreadsheet LCCA assessment is more than sufficient.

LIFE CYCLE COSTING AND ITS USES

Life cycle costing is in fact an outgrowth of several standard methods for calculating the economic impacts of various decisions over time. It is similar in many ways to the standard business measure of net present value (NPV), which takes future cash flows from any proposed investment (both in and out) and calculates their value in current dollars; when the outcome is positive, the project makes economic sense. A similar measure is the internal rate of return (IRR), which provides a percentage return on the investment after calculating costs and benefits from today forward; usually, companies have a hurdle rate, and an IRR must exceed that number for the project to be considered.

LCCA is in the same mould. It reduces energy project complexity down to one number that can then be easily compared with the same calculations performed on other projects. Because life cycle costing includes both current and future benefits and costs, and discounts for the time value of money, it gives a far more realistic picture than the payback period discussed earlier.

Table 4 Lighting retrofit example, Stage 1 (Current situation) and Stage 2 (Proposed situation)

	Stage 1 (current situation)	**Stage 2 (proposed situation)**
Description	100 fixtures; 2 bulb 4 ft long 34-Watt T12 FL fixtures with magnetic ballasts, ½ lenses broken, fixtures need cleaning but otherwise OK, fixture wattage 84, personnel unhappy with light provided; fixtures 2 years old	T8 lamps and electronic ballasts, to be installed in the same fixture; all fixtures to be cleaned. Total wattage 59, increased light levels 10%. Group replacement anticipated every 3 years
First costs		
Design cost		$1000
Lenses for fixtures	50 lenses at $8 each	50 lenses at $8 each
Ballasts		$10.50 each, group replacement
Bulbs		$3.00 each
Installation and cleaning cost (incl. labor)	$20.00 per fixture on "ad hoc" basis	$10.00
Fixture cost	$44.52	$44.52
Numbers of fixtures	100 (100 ballasts, 200 bulbs)	100 (100 ballasts, 200 bulbs)
Operating and maintenance costs		
Bulb replacement (incl. cleaning and labor)	$28 each, replaced as needed, average ¼/year	
Ballast replacement (incl. cleaning and labor)	$21.50 each, replaced as needed, average ¼/year	
Repairs	$200/year	
Electricity	$0.08 blended rate/kWh	
Replacement/retirement cost		
Expected life	4 years	4 years
End of life disposal costs—lamp	$1	$1
End of life disposal costs—ballast	$1.75	
Salvage value at end of life	0	0
General assumptions and costs		
Finance charges	0	0
Expected fixture life	10 years	10 years
Discount rate (nominal rate for 10 years)	6.1%	6.1%
Utility rate changes expected	Increase 2% yearly	Increase 2% yearly
Usage changes expected	None	None
Non-monetary costs	Complaints regarding lighting levels and color, "look" of the facility	
Non-monetary benefits		Increased light levels and better attractiveness of facility
Cost escalation projection	None	None
Hours of operation	24 h	24 h

Source: From Current-C Energy Systems, Inc., and Financial Energy Management, Inc.

Inv–Light

Life cycle costing can be used in different ways, depending on the situation and the circumstances:

- With new buildings (and principally through computer models incorporating life cycle costing), design, systems, site placement, materials, equipment choice, and other options can be modeled and evaluated so that the best possible choices are made, both in construction and in operating cost.
- With existing buildings, LCCA can be used to:
  - — Compare the current situation with that proposed under several design scenarios.
  - — Choose the optimum balance between retrofit cost and operating expenses.
  - — Choose one or several possible energy efficiency options among a number of possibilities.
- For states and other governmental agencies, as well as large corporations, LCCA models can be required for both projects and systems as prerequisites to approval, and often, the particular variables to be used and the specific computer model that is acceptable will be specified.

Many of the computerized technical and economic modeling programs are excellent, accounting for the interactions among various systems and possibilities, and giving a true picture of a complicated set of interrelationships. They are not fundamentally different, however, from a spreadsheet-based LCCA, which any manager can use with relative ease. Where a full economic analysis and computerized modeling for various design alternatives are required, they should be done by engineers or designers experienced in the particular modeling program to be used; their analyses of the competing alternatives and recommendations regarding options to install will very likely be worth far more than the cost of their services.

EXAMPLE

Let's take an example: a small lighting retrofit. First, Table 4 describes both the current situation and the proposed retrofit throughout the anticipated life of the project.

For the third stage of quantifying costs and benefits, additional variables are defined, and the LCCA spreadsheet is constructed for the life of the project—in this case, 8 years, which is the remaining projected life of the fixtures. The retrofit option also includes group relamping of bulbs and ballasts every 3 years, with a guarantee from the manufacturer to cover any materials and maintenance costs between those times (Tables 5 and 6).

Table 5 Lighting retrofit example, Stage 3 (Determine and quantify costs and benefits) – Current situation

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
First cost								
Design cost								
Lenses	$8×50							
Ballasts								
Bulbs								
Installation & cleaning incl. labor								
Fixture cost								
No. of fixtures								
O&M costs								
Total bulb replacement	$21.50 per bulb, 25% of 200 per year							
Total ballast replacement	$28.00×25%×100							
Repairs	$200	$200	$200	$200	$200	$200	$200	$200
Electricity	Fixture watts×number of fixtures×hours used in a year/1000, increasing 2%/year compounded							
Replacement/retirement cost								
Expected life	10 years total for fixtures; 8 years remaining							
Lamp disposal	$1 per lamp, 25% of 200 per year							
Ballast disposal	$1.75 per ballast, 25% of 100 ballasts per year							
Salvage value	$0							

Source: From Current-C Energy Systems, Inc., and Financial Energy Management, Inc.

Inv–Light

Table 6 Lighting retrofit example, Stage 3 (Determine and quantify costs and benefits) – Proposed situation

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
First cost								
Design cost	1000							
Lenses	$8×50							
Ballasts	$10.50×100			$10.50×100			$10.50×100	
Bulbs	$3×200			$3×200			$3×200	
Installation & cleaning incl. labor	$10×100			$10×100			$10×100	
Fixture cost								
No. of fixtures	100							
O&M costs								
Total bulb replacement								
Total ballast replacement								
Repairs								
Electricity	Fixture watts×number of fixtures×hours used in a year/1000, increasing 2%/year compounded							
Replacement/ retirement cost								
Expected life								
Lamp disposal	$1.00×100			$1.00×100			$1.00×100	
Ballast disposal								
Salvage value								

Source: From Current-C Energy Systems, Inc., and Financial Energy Management, Inc.

When the data has been input into the spreadsheet, Microsoft Excel or a similar program is used to calculate the life cycle cost of the proposed project based on that information.

Based on the assumptions shown above and with the cost and benefit calculations assigned to their particular years, cash flows for each year are first summed for each year and then discounted back to the present. In Excel, this can be done by using the NPV formula, which allows both discount rates and differential cash flow by period. If an organization does not have a generally accepted internal discount rate, appropriate rates for the United States, drawn from the President's annual budget submission, can be found in the ASTM International standard.[3]

Then the life cycle cost for the proposed project is compared with the life cycle cost for the current situation, and a decision can be made to proceed or not, based on a good economic assessment.

LIFE CYCLE COSTING CALCULATORS AND MODELS

The same approach and components could be used for widely different types of projects, and in fact, many LCCA calculators for an incredible variety of products are available on the Web sites of many organizations that are interested in fostering energy efficiency. Following are two examples:

- For an electric griddle, www.fishnick.com/tools/calculators/egridcalc.php is a calculator from the Food Service Technology Center, which also offers other food service-related calculators.[4]
- For transformers, the Copper Development Institute has an excellent set of formulas that can be used to evaluate alternative transformer possibilities. These formulas are available at www.copper.org/applications/electrical/energy/trans_life_cycle.html.[5]

As noted, a multitude of computer models include energy project LCCA in addition to their core focus areas (which may be energy costs, integrated building design, or other specialties). These are well summarized by the National Institute of Building Sciences[6] and are listed here for reference:

- Screening tools: FRESA, FEDS
- Architectural design tools: ENERGY-10, Building Design Advisor, Energy Scheming

- Load calculation and HVAC sizing tools: Hap, TRACE, DOE-2, BLAST, VisualDOE, EnergyPlus
- Economic assessment tools: BLCC, Quick BLCC.

CONCLUSION

Life cycle costing analysis is a tool that combines financial, technical, and other information considered over time to facilitate decision-making. The variables in use, the cost of whole projects or systems, and regulatory requirements can make the use of complex computer modeling vital. The principles of LCCA, however, are relatively simple and straightforward. The techniques can be used for small as well as large decisions, and provide managers and decision-makers a way to compare possibilities that are not on the surface comparable.

Mastering the techniques of simple LCCA calculation is possible for any well-informed professional. Such mastery in turn facilitates communication among those professionals and technically trained experts needed for more complex modeling. Thus, LCCA can be a useful tool both for making decisions and for talking about those decisions.

ACKNOWLEDGMENTS

The author would like to thank the following individuals and organizations that have assisted in the preparation of this document (without them, the pages would have long remained blank): Financial Energy Management, Denver, Colorado, for its assistance in the lighting example; Don Croy, PE, at Current-c Energy Systems, for other examples; and Lorrie Tietze and Dr Neal Mangham for their enthusiasm and editing.

REFERENCES

1. Fuller, S. Life Cycle Cost Analysis. Available at: www.wbdg.org/design/lcca.php (accessed on November 15, 2005).
2. Available at: www.focusonenergy.com/2062%20SAWlife-cycle.pdf (accessed on November 15, 2005).
3. ASTM International E917-05 Standard Practice for Measuring Life-Cycle Costs of Buildings and Building Systems. Available at:www.astm.org/cgi-bin/SoftCart.exe/DATA-BASE.CART/REDLINE_PAGES/E917.htm?L+mystore+uskj1629+1132110605, Appendix C.6 (accessed on November 15, 2005).
4. Available at: www.fishnick.com/tools/calculators/egridcalc.php (accessed on November 15, 2005)
5. Available at: www.copper.org/applications/electrical/energy/trans_life_cycle.html (accessed on November 15, 2005)
6. National Institute of Building Sciences Whole Building Design Guide. Available at: www.wbdg.org/design/energyanalysis.php (accessed on November 15, 2005).

BIBLIOGRAPHY

1. ASTM International (www.astm.org).
2. Copper Development Institute (www.copper.org).
3. Energy Star (www.energystar.gov).
4. Food Service Technology Center (www.fishnick.com/tools/calculators/egridcalc.php).
5. National Institute of Building Sciences Whole Building Design Guide (www.wbdg.org/design/energyanalysis.php).
6. National Institute of Standards and Technology (NIST) (www.nist.gov).
7. www.pumps.org/public/pump_resources/energy/home.htm).

Lighting Controls

Craig DiLouie
Lighting Controls Association, National Electrical Manufacturers Association, Rosslyn, Virginia, U.S.A.

Abstract
Automatic lighting controls have become a standard feature in new construction due to prevailing energy codes. Advancements in technology now enable a broad range of globalized and local, automatic and manual control strategies to generate energy savings and support visual needs.

INTRODUCTION

Lighting controls are an essential part of every lighting system. The most basic example of a lighting control is the common switch. When the switch is flipped on, the circuit is closed, enabling the flow of current to operate the connected lighting system.

A wide variety of devices and systems are used to control lighting systems in commercial and industrial buildings. Of particular interest is lighting automation, which presently represents a major frontier in building and energy management. Automated lighting controls automatically switch or dim lighting systems based on a given input. The automation of this functionality can provide significant benefits.

In some industries, lighting accounts for more than 60% of a facility's electrical bill and 40% of the total energy bill. Automated lighting controls can contribute significantly to operating cost savings by switching or dimming the lights according to time of day, whether the space is occupied, amount of available daylight, current level of light output from the lamp, and other factors. According to the New Buildings Institute, which developed the *2001 Advanced Lighting Guidelines*, automatic lighting controls can reduce lighting energy consumption by 50% in existing buildings and at least 35% in new construction. Related to a reduction in energy consumption are load shedding and peak demand reduction, which can reduce demand charges imposed by the utility (Fig. 1).

In addition, lighting automation can provide mood setting, via the ability to alter a space through dimming or color changing; flexibility, by allowing users to instantly adapt a space to different uses; the ability to establish a responsive lighting system that can be globally and locally controlled with automatic operation; the ability to adapt electric lighting systems to daylighting strategies; enhanced security; decreased "light pollution" (skyglow, light trespass, and glare), by dimming or switching outdoor lights based on time of night or occupancy; enhancement of workspaces with a technology that has visible effects; and potential increased worker satisfaction, by enabling users to control their own light levels. The list goes on.

Lighting automation can be completely automated or can contain elements of manual operation; it can be localized, global or both; it can be hardwired or wireless; and it can be used for switching or dimming. A wide variety of proven and developing technologies is now available to achieve a wide variety of building and energy management goals (Table 1).

COMMON STRATEGIES

Lighting controls can perform one or more of seven basic functions: on/off, occupancy recognition, scheduling, task tuning, daylight harvesting, lumen depreciation compensation, and demand control.

On/Off is the basic control function, achieved through switching.

Occupancy recognition is used in intermittently occupied areas, typically to turn lights on automatically when the area becomes occupied and off when it becomes unoccupied. Some occupancy recognition devices are manual-on, automatic-off. The occupancy sensor technology is typically ultrasonic (sensing changes in transmitted sound waves returning to the device), passive infrared (sensing changes in heat in the area), or a combination of the two.

Scheduling is used for areas of predictable occupancy, in which a control system dims, activates or shuts off a lighting system on a predetermined schedule. Local manual overrides are usually provided. A typical weekday lighting schedule is shown in Fig. 3.

Tuning entails adjusting the light output of a lighting system to a desired level needed for a task or other purpose, such as aesthetics or mood setting. This can be achieved through either dimming or switching layers of lighting (such as bi-level switching, in which half of a lighting system is shut off in a space while the other half continues operating).

Keywords: Lighting; Controls; Switching; Dimming; Lamps; Ballasts; Occupancy sensors; Scheduling; Daylight harvesting.

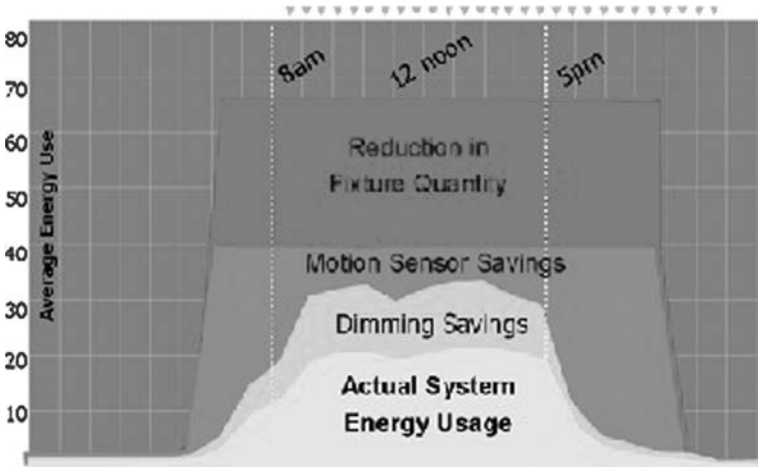


Fig. 1 A complete lighting upgrade can save energy on multiple levels, including fixed load reduction through reduced fixtures or lower-wattage lamp ballast systems, and variable load reduction through strategies such as occupancy recognition, scheduling and personal dimming.
Source: From Light Right Consortium.

Daylight harvesting is used to enable lighting systems to respond to ample available daylight by dimming or some level of switching.

Lumen depreciation compensation takes advantage of the fact that most lighting systems are overdesigned to account for a gradual reduction of light output of lamps as they advance in operating age. Similar to daylight harvesting, this strategy entails measuring available light in the space and dimming the light output to maintain a constant level.

Demand control involves switching or dimming lighting systems during utility peak demand periods, which results in energy cost savings and also potentially significantly lowers demand charges (Tables 2 and 3).

TYPICAL CONTROLS

Lighting controls generally can be categorized as manual or automatic, although some may have features of both

Fig. 2 A growing number of light fixtures are now available with integrated automatic lighting controls, such as this recessed basket, which includes an integrated sensor.
Source: From Lightolier.

Table 1 Benefits of automatic lighting controls in various space types

Space type	Benefit
Discount retail store	In an open retail space with daylighting, dimming can reduce electric lighting use but allow the lights to be on, making the store seem "open for business"
Conference room, classroom, auditorium, etc.	Dimming lighting can facilitate a variety of visual presentations
Health care facility	Daylight-driven dimming can provide a smooth and unnoticeable transition to electric lighting as daylight levels decrease, while maintaining the desired light level
Restaurant	Preset scene dimming controls can make changing the ambiance as the day goes on consistent and as easy as pressing a button
Office area	Even in an open office area, occupants can be given the option of dimming the light fixture over their workstation to suit their personal preferences

Source: From Lighting Controls Association (see Ref. 1).

types of operation. Manual controls require immediate human intent to turn the lighting on or off or adjust its output. Automatic controls initiate actions for lighting systems based on registered events or programming.

Control options can be grouped as switching controls, dimming controls, and integrated lighting control systems.

Switching controls turn lights on and off, and many perform other functions as well. At a minimum, every space should be equipped with manual switching to permit occupants or facility operators to control lighting usage.

Lighting contactors: lighting contactors permit manual or automatic control of large blocks of lighting loads.

Local wall switches: local wall switches (a.c. snap switches) are the most commonly used control devices for local lighting control. For best results, switches should be located to be convenient to users and to encourage deactivation of lighting whenever appropriate. Wall switches also can be applied to develop a flexible lighting control scheme by layering the lighting in the space, with different layers controlled by different switches.

Key-activated switches: key-activated switches are wall switches that turn lighting on and off by a key. They are installed to prevent unauthorized or accidental use of certain lighting circuits. They are particularly useful for HID light sources that must cool down before they can be activated.

Intelligent on/off local devices: intelligent on/off local devices consist of at least two elements: (1) a logic or intelligence module, and (2) a power switching device. The logic or intelligence elements vary depending on the needs of the specific applications. The intelligent input in its simplest form can be a time control or an occupancy sensor. Each typically is used to control a single load and is wired directly to it.

Time controls, also known as time clocks or time switches, activate and deactivate their loads at user-determined times, and are therefore most suitable for areas where occupancy is predictable (with local overrides often being desirable). Some are microprocessor based, allowing operators to program more on/off actuations per day, as well as create special schedules for holidays or certain functions. Outdoor lighting time controls often use an astronomical feature that automatically compensates the control schedule for sunrise and sunset time shift during the year.

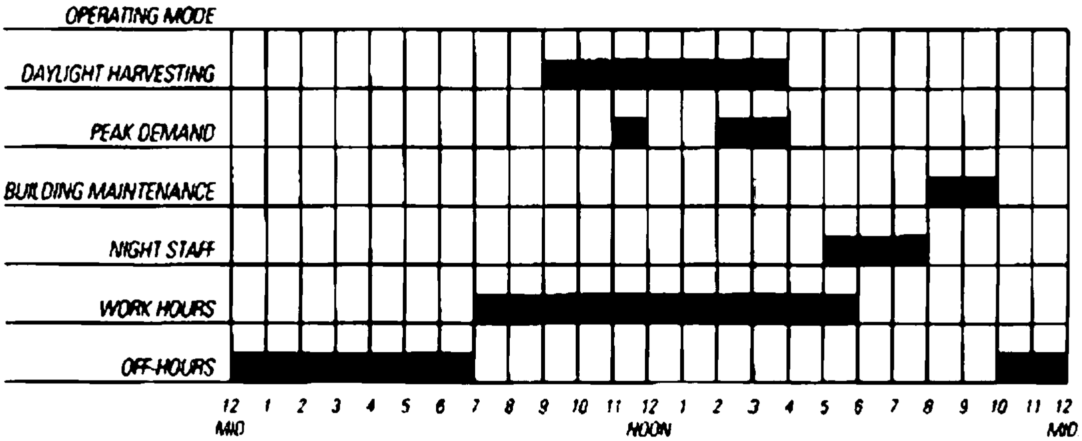


Fig. 3 Typical weekday lighting schedule.
Source: From Lighting Controls Association (see Ref. 2).

Table 2 Typical lighting control applications

Typical lighting control applications			
Type of control	**Private office**	**Open office—daylit**	**Open office—interior**
Occupancy sensors	++	++	++
Time scheduling	+	++	++
Daylight dimming	++	++	0
Bi-level switching	++	+	+
Demand lighting	+	++	++

++, good savings potential; +, some savings potential; 0, not applicable.
Source: From Federal Energy Management Program, U.S. Department of Energy (see Ref. 3).

Occupancy sensors are automatic switches that control lighting based on the presence or absence of people (see Figs. 5 and 6). Their primary function is to switch electric illumination off automatically in an unoccupied space after the last person leaves that space, saving energy (see Table 4). A timing control provides light for a period of time after the area is vacated.

Table 4 Typical energy savings with occupancy sensors

Occupancy area	Energy savings (%)
Private office	13–50
Classroom	40–46
Conference room	22–65
Restrooms	30–90
Corridors	30–80
Storage areas	45–80

Source: From U.S. Environmental Protection Agency.

Photocell controls respond to changes in ambient light. When the ambient light level falls to a user-determined level, lighting is switched on. When the ambient light increases to a user-determined level, lighting is switched off.

Low-voltage controls: low-voltage switching systems provide a more flexible switching platform than standard line-voltage switches. The simplest system consists of a transformer that produces 24 V or less, relays that are wired to the loads, and on/off switches that are connected by low-voltage wiring to the relays. Each relay can control up to one full branch circuit (20 A).

Table 3 Operating cost comparisons for private office and open office spaces, using various types of controls

Operating cost comparison private office, 128 ft^2

Performance	Base case	Occupancy sensors	Daylighting	Occupancy sensor + daylighting
Annual energy use[a]	450 kWh	340 kWh	330 kWh	250 kWh
Annual energy cost	$33	$24	$24	$18
Annual energy cost savings	—	$9	$9	$15

Operating cost comparison open office area, 1000 ft^2

Performance	Base case	Time scheduling	Occupancy sensors	Daylighting	Time scheduling + daylighting
Annual energy use[b]	5700 kWh	5100 kWh	5000 kWh	4200 kWh	3700 kWh
Annual energy cost	$340	$305	$300	$250	$220
Annual energy cost savings	—	$35	$40	$90	$120

[a] Average daily "on" hours for wall switch is 14.7. Average daily occupied hours for the office is 12.9.
[b] Average daily "on" hours for wall switch is 9.1. Average daily occupied hours for the office is 6.8., Cost-effectiveness assumptions: Each of the two operating cost comparisons assumes that the workspace has approximately 1.5 W/ft.^{2} of ceiling lighting, with parabolic troffer luminaires containing T-8 lamps and electronic ballasts. Daylighting examples assume a design light level of 55 footcandles at work surfaces. Assumed electricity price: $0.06/kWh, the Federal average electricity price (including demand charges) in the U.S.

Source: From Federal Energy Management Program, U.S. Department of Energy (see Ref. 3).

Fig. 4 Lighting control panel installation.
Source: From HUNT Dimming.

Fig. 5 Occupancy sensors.
Source: From Leviton Manufacturing.

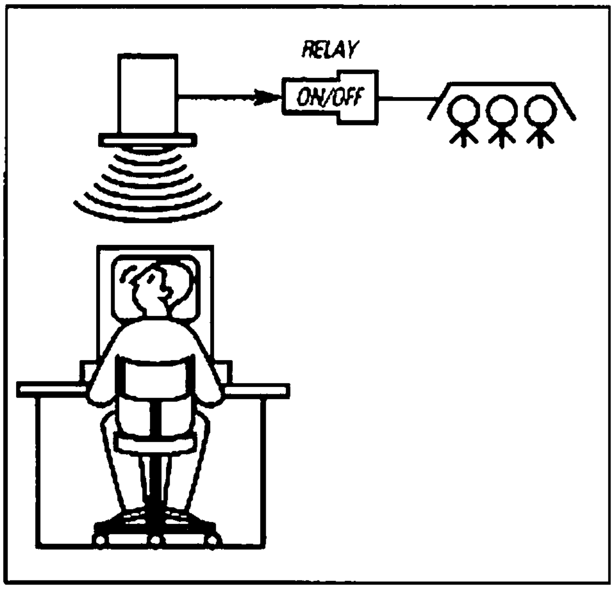


Fig. 6 Occupancy sensor system.
Source: From Lighting Controls Association (see Ref. 2).

Low-voltage wiring provides inherent wiring flexibility while also providing the foundation for simple lighting automation (Fig. 7).

Step-level HID controls: step-level HID lighting controls are relay systems that operate mercury vapor, metal halide, and high-pressure sodium lighting at either full light output or less (e.g., 50%) (Fig. 8).

Dimming controls are available for most types of lighting. They can be integrated into automatic lighting control systems and can be used manually as well. Some dimming controls require use of dimming ballasts, while others employ an electronics package installed in the panelboard or elsewhere within the system.

Wallbox dimmers: wallbox dimmers are manual controls that give occupants more control over their visual environment. Various control configurations are available, including linear slides, rotary knobs, raise/lower buttons, preset panels, and wireless remotes.

Integrated dimmers: integrated dimmers integrate a variety of features into a wallbox configuration. Commonly included features are multiple channel control, with which all or selected fixtures on a circuit are controlled by a single dimmer; multiple presets; and universal circuitry that allows each dimming channel to control incandescent, low-voltage incandescent, fluorescent, cold-cathode, or neon light sources (Fig. 9).

System dimmers: system dimmers offer lighting control for larger applications in which wallbox and integrated products are impractical or higher performance is required. These systems consist of dimmer cabinets and control stations, typically connected with low-voltage control wires (Fig. 10).

Dimming ballasts: variable output ballasts integrate dimming capability into fluorescent (see Fig. 11) and HID (see Fig. 12) ballasts so that they can dim the lamps according to manual, scheduled, or event-based input. An example of manual input is giving each occupant personal dimming control of his or her lighting system. An example of scheduled input is dimming the lamps across areas of a facility at preset times of day for demand control. An example of event-based input is daylight harvesting, in which a photocell is used to measure light levels, and the ballast dims the lamps to maintain a preset light level based on whether ample daylight is available.

Programmed-start ballasts: most available dimming ballasts for fluorescent lamps are rapid-start electronic ballasts or programmed-start electronic ballasts. Programmed-start ballasts are of interest with or without dimming capability. They are rapid-start ballasts that

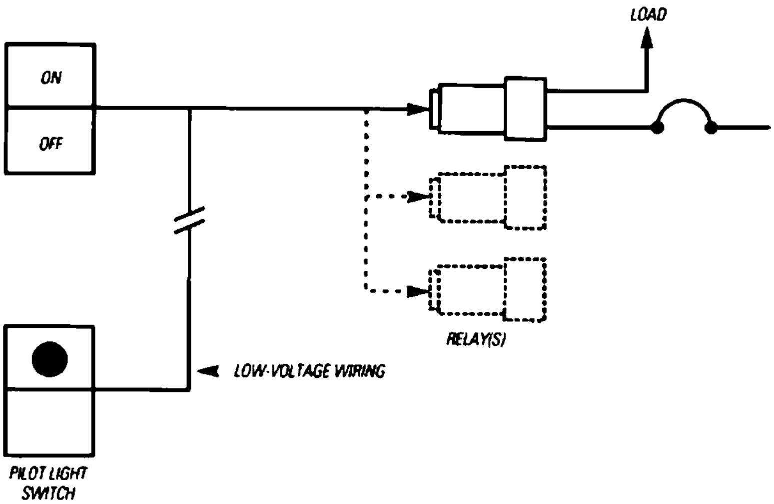


Fig. 7 Low-voltage control system schematic.
Source: From Lighting Controls Association (see Ref. 2).

Inv–Light

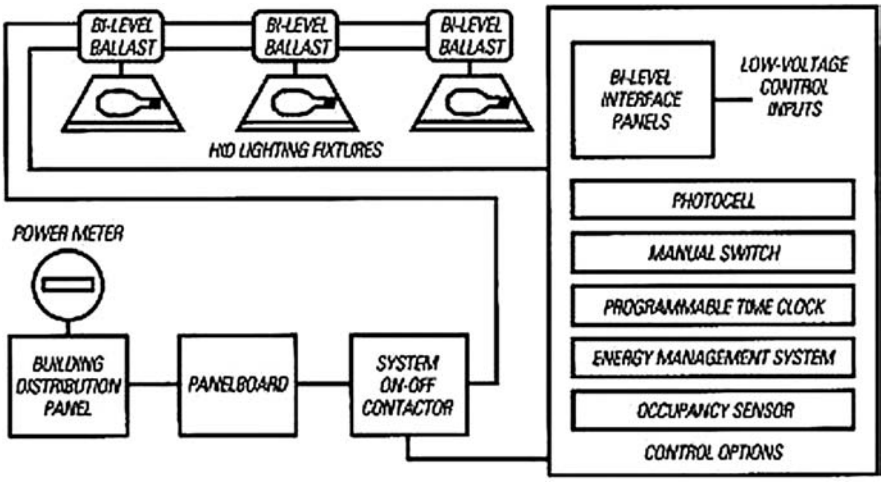


Fig. 8 Schematic of two-level HID lighting control system.
Source: From Lighting Controls Association (see Ref. 2).

preheat the electrodes more accurately to minimize damage to the electrodes during the startup process (according to a program), and therefore can optimize lamp life. While supplying the preheat voltage, the ballast minimizes the lamp voltage, thereby reducing glow current during this phase with its associated degrading effect on lamp life. As a result, programmed-start ballasts can provide up to 100,000 starts, ideal for applications where the lamps are frequently switched, such as spaces with occupancy sensors.

Integrated lighting control systems consist of manual or automatic components designed to control compatible

Fig. 9 Dimming system.
Source: From Leviton Manufacturing.

Inv–Light

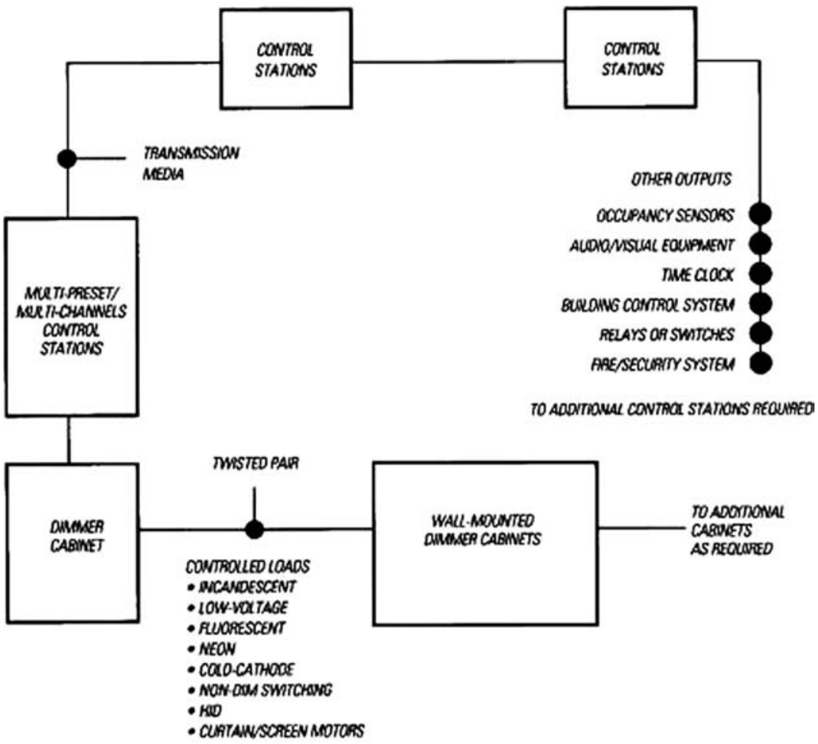


Fig. 10 System dimmer.
Source: From Lighting Controls Association (see Ref. 2).

dimming ballasts. They can be interfaced with other microprocessor-based centralized lighting control systems or building automation systems. These systems can perform all of the functions that are important to energy optimization. They can sense conditions in each area or zone and control lighting to yield maximum energy efficiency without affecting visual comfort or other conditions (Fig. 13).

Microprocessor-based centralized programmable lighting control: a microprocessor-based centralized programmable lighting control system is basically a microprocessor-based centralized controller. Although it

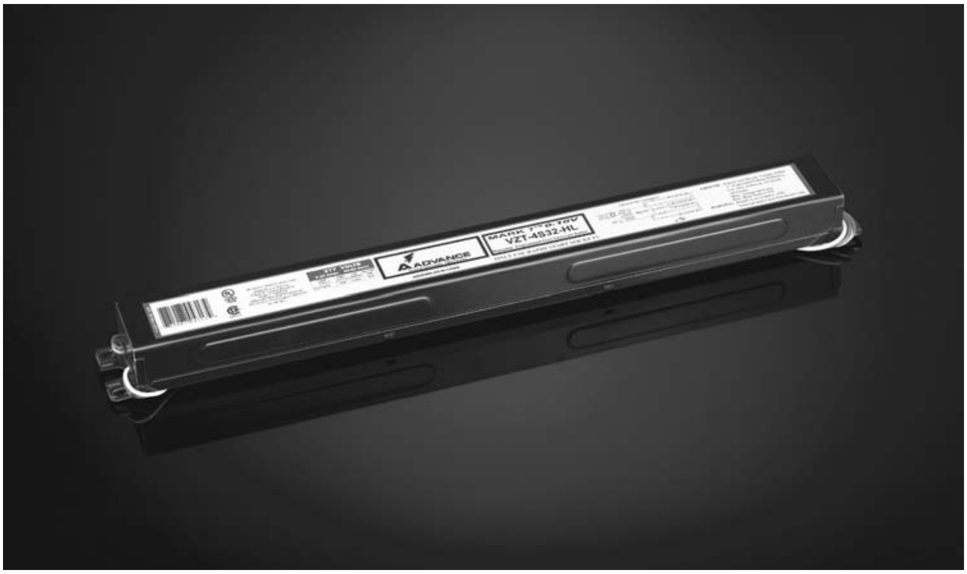


Fig. 11 Fluorescent dimming ballast.
Source: From Advance.

Fig. 12 HID dimming ballast. HID dimming ballasts are relatively new and are growing in popularity.
Source: From Advance.

is designed principally for lighting, it is capable of handling other loads. Photocells and other controls can be integrated into the system.

Microprocessor-based programmable controllers can be integrated into networked lighting control systems that allow schedules and other programmable functions to be entered and then changed from a central operator console. Networked systems also allow input from devices such as master switches, photocells, occupancy sensors, telephones, or load-shed contacts to control relays or dimmers. In addition, the network allows the central collection of operating data and status information.

The typical networked control system shown in Fig. 14 provides automated lighting control for applications ranging from a small office building to a mall to an industrial complex. Each of the distributed control panels has standalone automation capability. The network links these controllers to a central operator terminal (PC). Besides supporting such features as telephone control and distributed master switching, these systems can provide energy management data. Providing networked lighting controls also ensures that the lighting can be effectively integrated with other building controls (controlling security, etc.) to provide full intelligent lighting operation.

Building automation systems: building automation controls generally are microcomputer- or minicomputer-based systems that are capable of controlling lighting systems as well as HVAC, security, and fire safety systems.

EVALUATING LIGHTING CONTROL OPTIONS

Once the general type of control system has been decided upon, the next step is specifying the specific control devices. In order to do so, it is essential to establish criteria for evaluating options. Some of the key criteria to consider are given below.

Cost-effectiveness: lighting controls that reduce energy consumption and/or reduce demand are an investment in profitability. Control options should be evaluated based on initial installed cost vs operating cost savings to determine return on investment and the payback period. Options that generate the highest return on investment and meet other selection criteria are more desirable.

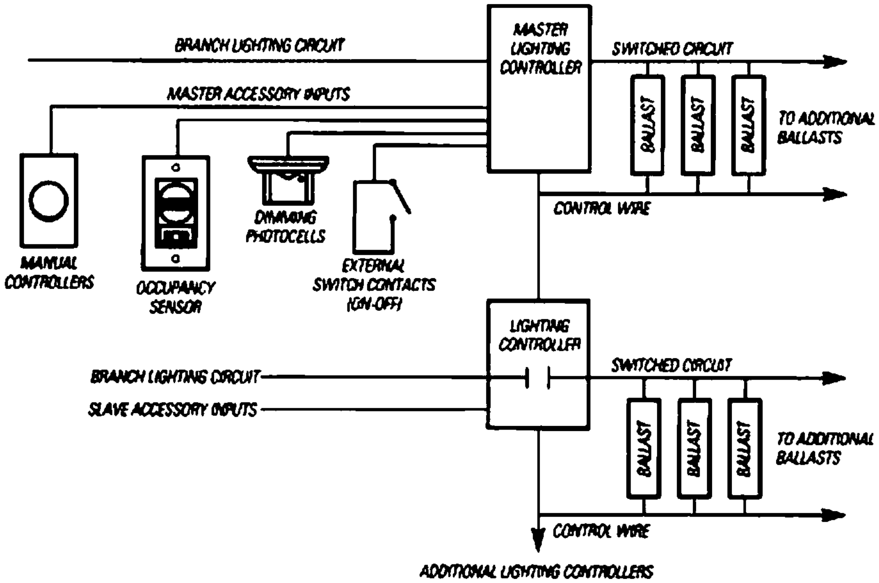


Fig. 13 Integrated lighting control system.
Source: From Lighting Controls Association (see Ref. 2).

Inv–Light

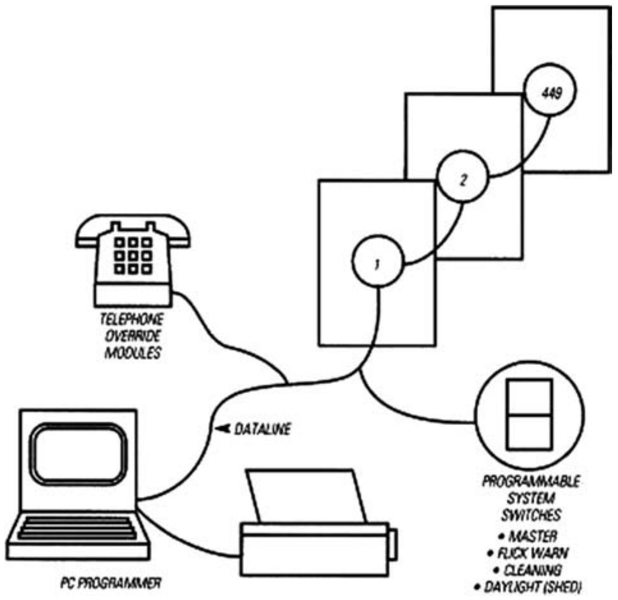


Fig. 14 Networked intelligent panels.
Source: From Lighting Controls Association (see Ref. 2).

Adaptability: for projects in existing facilities, lighting controls should be evaluated based on how well they can be adapted to the facility, and how advantageous they will be. For example, can the new control system be interfaced existing local controls, or will the controls have to be replaced? If daylight harvesting is of interest, is there sufficient daylight to warrant it? Adaptability applies not only to applying controls to existing spaces; but also to ensuring, in both existing spaces and new construction, that the control system is scalable to future space use and potential advances in technology.

Flexibility: controls can be evaluated based on the degree of flexibility required from the lighting system, which in turn is based on how the space is used. In a gymnasium, for example, a variety of activities can occur, from sporting events to assemblies and plays. A flexible lighting system enabled by appropriate controls can respond to all of these uses.

Maintainability: maintainability refers to the ease with which a system is maintained, something determined through evaluation of two principal factors: (1) in-house maintenance support through training programs and manuals, and (2) the availability of professionally trained maintenance persons employed, licensed, or authorized by the manufacturer.

Wiring diagrams of the system also are essential during maintenance procedures; someone who understands the diagrams must be available. Do not assume that a given manufacturer can always provide effective maintenance on a timely basis. Check references to determine how satisfied other users of the proposed system are with the maintenance services provided. Key concerns in this regard are completeness of preventive maintenance, responsiveness and capability of outside maintenance, and the availability of a service maintenance agreement.

Reliability: reliability relates to how well the system performs and the way in which it performs. System performance can be determined primarily by talking with other users of the system. They can relate how often breakdowns occur and the time required to restore equipment to its pre-failure condition. Most modern controls are highly reliable.

Programmability: programmability is the degree to which the programming capability of a microprocessor-based lighting control system can be modified. Many programs are written in such a way that they can be modified relatively easily, either by the user or the manufacturer.

Complexity: lighting control options can be evaluated by their ease of specification, installation, programming, commissioning, and ongoing use.

Interoperability: all components of a desired control system, which may include components as diverse as relays, photocells, dimmers, dimming ballasts, and software, must be electrically compatible and operate reliably properly as a system. This may require consideration of control voltage and also operating protocol such as 0-10VDC, DALI, and BACnet.

APPLICATION ISSUES

Design controls for each area: the first step is to systematically evaluate all the parameters involved in the design in light of the design goals. For each area, the designer will need to determine which components will be most appropriate. At the same time, the designer will need to decide on the optimum placement for each component. Often, the control manufacturer can offer valuable assistance for each phase of the project, from evaluation to providing wiring diagrams to give to the contractor.

Zoning for occupancy recognition: occupancy recognition controls can improve usability, security, and efficiency in a building. However, they can be bypassed or even removed by the owner if not applied properly. The designer should take into account how the space is used, including how it's occupied. For example, if an employee shows up on a Saturday, the entire floor's lighting should not activate, only the local areas the employee is using. Similarly, if the employee is sitting at his or her desk, the lights should stay on by continually detecting their occupancy.

Prevailing energy codes: require the use of automatic lighting controls. The U.S. Department of Energy, via a ruling under the Energy Policy Act of 1992, now requires that all states have an energy code at least as stringent as ASHRAE/IES Standard 90.1-1999.

Table 5 Selection of controls for various types of spaces: room by room analysis

Space type	Use pattern	IF…	THEN…
Cafeterias or lunchrooms	Occupied occasionally	Daylighted…	Consider daylight-driven dimming or on/off control
		Occupied occasionally	Consider ceiling-mounted occupancy sensor(s). Make sure minor motion will be detected in all desired locations
Class room	Usually occupied occasionally occupied	Multi-tasks like overhead projectors, chalkboard, student note taking and reading, class demonstrations	Consider manual dimming
		Occupied by different students and teachers	Consider ceiling- or wall-mounted occupancy sensor(s) and manual dimming. Make sure that minor motion will be detected
		Lights left on after hours	Consider centralized controls and/or occupancy sensors
Computer room	Usually unoccupied	Lights are left on all the time	Consider occupancy sensors with manual dimming. Be sure that minor motion will be detected and that equipment vibration will not falsely trigger the sensor
Conference room	Occupied occasionally	Multi-tasks from video-conferencing to presentations	Consider manual dimming (possibly preset scene control)
		Small conference room	Consider a wall box occupancy sensor
		Large conference room	Consider ceiling- or wall-mounted occupancy sensor(s). Be sure that minor motion will be detected in all desired locations
Gymnasium or fitness	Usually occupied	Requires varied lighting levels for activities	Consider manual dimming and occupancy sensors. Be sure that the HVAC system will not falsely trigger the sensor
Hallways	Any	Occasionally or usually occupied	Consider occupancy sensors with elongated throw. Be sure that coverage does not extend beyond the desired area
		Daylighted…	Consider daylight on/off control
Health care—examination rooms	Occasionally occupied	Different lighting needs for examination	Consider manual dimming
		Small areas	Consider a wall box occupancy sensor
Health care—hallways	Usually occupied	Daylighted…	Consider automatic daylight-driven dimming
		Requires lower lighting level at night	Consider centralized controls to lower lighting levels at night
Health care—patient rooms	Usually occupied	Different lighting needs for watching television, reading, sleeping and examination	Consider manual dimming. Occupancy sensors may not be appropriate
Hotel rooms	Occasionally occupied	Use primarily in the late afternoon through evening for sleeping and relaxing	Consider manual dimming
Laboratories	Usually occupied	Daylighted…	Consider automatic daylight-driven dimming in combination with occupancy sensors

(Continued)

Table 5 Selection of controls for various types of spaces: room by room analysis *(Continued)*

Space type	Use pattern	IF…	THEN…
Laundry rooms	Occasionally occupied	Requires high light levels, yet lights are usually left on	Consider occupancy sensors
Libraries—reading areas	Usually occupied	Daylight…	Consider automatic daylight-driven dimming
		Lights left on after hours	Consider centralized controls
Libraries—stack areas	Occasionally occupied	Stacks are usually unoccupied	Consider ceiling-mounted sensor(s)
Lobby or atrium	Usually occupied but no one "owns" the space	Daylighted and lights should always appear on…	Consider automatic daylight-driven dimming
		It isn't a problem if lights go completely off in high daylight…	Consider automatic daylight-driven dimming or on/off control
		Lights are left on all night long, even when no one is in the area for long periods	Consider occupancy sensors. Be sure that minor motion will be detected in all desired areas
Office, open	Usually occupied	Daylighted…	Consider automatic daylight-driven dimming
		Varied tasks from computer usage to reading	Consider manual dimming
		Lights left on after hours	Consider centralized controls and/or occupancy sensors
Office, private	Primarily one person, coming and going	Daylighted…	Consider manual dimming, automatic daylight-driven dimming, or automatic on/off
		Occupants are likely to leave lights on and occupants would be in direct view of a wall box sensor	Consider a wall box occupancy sensor
		Occupants are likely to leave lights on and partitions or objects could hide an occupant from the sensor	Consider a ceiling- or wall-mounted occupancy sensor
Photocopying, sorting, assembling	Occasionally occupied	Lights are left on when they are not needed	Consider an occupancy sensor. Be sure that machine vibration will not falsely trigger the sensor
Restaurant	Usually occupied	Daylighted	Consider automatic daylight-driven dimming
		Requires different lighting levels throughout the day	Consider manual dimming (possibly preset scene dimming)
		Requires different lighting levels for cleaning	Consider centralized control
Restroom	Any	Has stalls	Consider a ceiling-mounted ultrasonic occupancy sensor for full coverage
		Single toilet (no partitions)	Consider a wall switch occupancy sensor
Retail store	Usually occupied	Daylighted…	Consider automatic daylight-driven dimming

(Continued)

		Different lighting needs for retail sales, stocking, cleaning	Consider centralized controls or preset scene dimming control
Warehouse	Aisles are usually unoccupied	Daylighted	Consider daylight-driven dimming or daylight on/off control
		Lights in an aisle can be turned off when the aisle is unoccupied	Consider ceiling-mounted occupancy sensors with elongated throw. Select a sensor that will not detect motion in neighboring aisles, even when shelves are lightly loaded

Source: From Lighting Controls Association (see Ref. 1).

Utility rebates and incentives: a number of utilities continue to offer financial incentives for organizations that upgrade their lighting systems. The incentive may take the form of cash for installing approved technologies or cash per removed unit of energy consumption or demand.

Electrical design: lighting control systems should be designed using all appropriate code rules and design practices related to overload, short-circuit protection, grounding, and other safety concerns. Be sure the control system can handle the steady-state current, lamp inrush, ballast harmonics, and available fault currents (which may entail tradeoffs among these factors). Do the electrical design before working out the details of the control scheme. For example, ensure that there is sufficient space in the electrical closet for needed components (Table 5).

Switching vs dimming: in daylight harvesting and similar schemes, switching lamps can be jarring if it occurs frequently during the day. Smooth dimming between light levels is less noticeable by occupants.

Layering control systems: a number of buildings utilize schedule-based systems as a global switching system and supplement them with occupancy sensors, manual switches, and dimming systems as needed. The schedule-based system activates and shuts off zones or the entire building based on predicted occupancy; the occupancy sensors and manual switches provide local override and automatic off after occupants leave; and dimming is used for architectural spaces, conference rooms, and areas of opportunity for daylight harvesting. Additional overrides can be provided by telephone or network interface for unusual circumstances. Zones should be sufficiently small for flexibility and maximum savings, although one should be aware of additional installed cost if zones are too small. Ensure that changing the schedule is easy, and that the system provides a flexible design that allows for different schedules for areas of the building with different needs and alternate schedules for weekends and holidays (Fig. 15).

Select products: depending on the relative importance of the several factors, select appropriate components and test the integrated design to see if it will satisfy the goals. Watch out for mismatched components. For fluorescent lighting, ballasts and controls must be compatible. Fluorescent fixtures that are intended to be dimmed require special dimming ballasts. There are several kinds of control systems, and likewise, there are several varieties of dimming ballasts. For the switching of outdoor lighting systems, a more robust device will likely be needed than those used indoors.

Control voltage: some controls can be connected to line-voltage power, while others must be connected to low-voltage (d.c.) power. Line-voltage controls tend to be less expensive, but less flexible, than low-voltage controls. They may also be more cost-effective for retrofits. In many cases, however, control systems require low-voltage

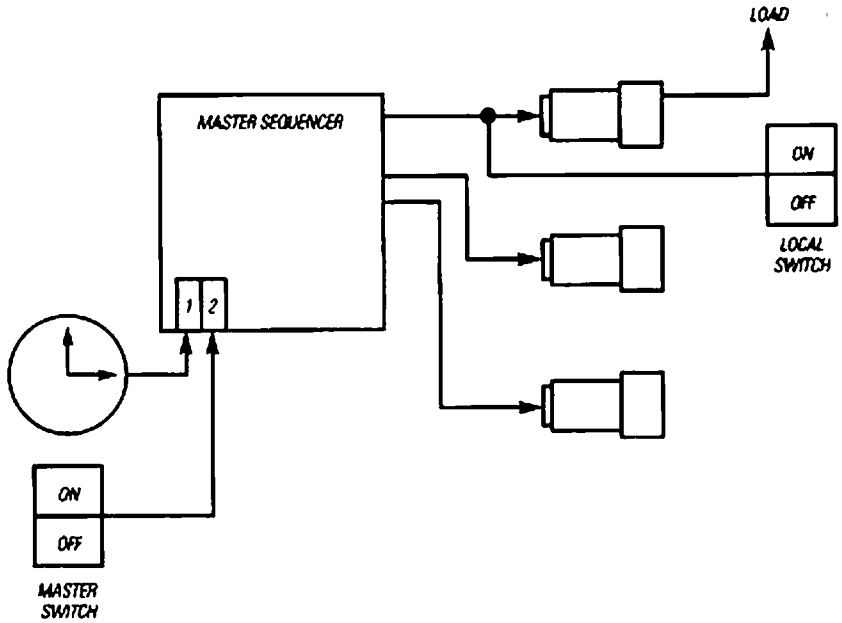


Fig. 15 Lighting automation with local switch override.
Source: From Lighting Controls Association (see Ref. 2).

control because low-voltage components such as light-level sensors are present.

Equipment location: as the designer locates the controls, he or she should ask if their placement is appropriate. Controls should be easy to locate and access for use or maintenance. Partitions and walls will affect the coverage patterns of sensors. For occupancy sensors, be aware of space considerations that could result in false triggering. The "brains" of lighting control systems should be mounted near the lighting panelboards in the electrical closets, requiring sufficient space (Table 6).

Typical coverage patterns (applies to occupancy sensors): there are several different kinds of coverage patterns and mounting configurations for occupancy sensors, such as ceiling-mounted controls with 360° coverage, ceiling-mounted controls with elongated "corridor" coverage, wall-mounted controls with a fan-shaped coverage pattern, and ceiling-mounted controls with a rectangular coverage pattern. Take note of the difference between each device's sensitivity to minor motion (working at a desk) vs major motion (walking or half-step activity). The manufacturer should provide coverage diagrams for both levels of activity.

Light levels: watch out for inadequate light levels. Make sure that set points are selected that will please the majority of the occupants. Get their input if possible. Follow IESNA recommendations based on the type of tasks performed in the area.

Ease of use and maintenance: the controlled lighting should be seen from the control panel or switch location. If the controls adapt to the normal behavior of people, they will be accepted by occupants. If not, they will be rejected. Similarly, if controls aren't simple, they will not be used. Controls should make sense and provide flexibility to all users.

Security issues: in high-security applications, occupancy sensors will indicate that people are present wherever lights are on. It is also advisable that, in these areas, there should be no manual-off option, and sensors should be protected from tampering and vandalism.

Construction documents: a complete set of construction documents includes, but is not necessarily limited to, (1) drawings, showing control locations, circuiting, and a control zone diagram to show which light fixtures are controlled by which device and how the controls are interrelated; (2) wiring diagrams for control components; (3) a schedule of controls, showing catalog numbers and descriptions of selected products (including all necessary power packs and accessories); and (4) written specifications for the control system, explaining the work and submittals included and clearly describing the approved equipment needed to achieve the desired results.

Installation: it pays to make sure the contractor understands the way the control scheme works. In one high-profile project, the light-level sensors were supposed to be installed underneath the indirect light fixtures. Instead, they were initially installed on top of them. When the sensors determined that more light was needed, they turned the lights on. But, when the lights came on, they shone on the sensors—so off they went again.

Commissioning: when the installation is complete, the controls should be commissioned, which often

Table 6 Questions to ask lighting control system vendors

Financial arrangements	What is the total installed cost of the system and what does it include? Over what time period must it be paid?
	Will the supplier provide certain services? What are they?
	Can the equipment be leased or rented? How much is it? What does the cost include (and exclude)?
	What are the delivery lead times?
Reliability	How many units of the model under consideration are currently installed?
	When was the first unit of this model installed?
	What is the term of warranty and what does it cover? Under what circumstances can the warranty be extended? Will the seller warrant against damage to any other purchaser's equipment? What are the general liability limits and how are claims settled?
Maintenance	How much does it cost and what does it cover?
	What schedules are available? (Most companies offer many schedules, depending on the response time and coverage desired.)
	Where is the nearest service office?
	Have they been trained on the piece of equipment being considered? (If this equipment is the only model of that type installed, chances are that a great deal of on-the-job training will be provided for in-house service people.)
	What spare parts are recommended?
	How many different modules does the system have?
	Does the company selling and installing the equipment also manufacture and service it?
	Is the system configured so that it can be backed up?
References	Who in the area is using this model equipment?
	How is it being used?
	Is the application comparable to the one being contemplated?
	How long has the equipment been in place?
Training	How much training is required?
	Is it included in the purchase price?
	How much does extra training cost?
	Where does training take place? How long will it last?
General	Is applications support generally required?
	Can the system be expanded or upgraded easily?
	How much will expansion cost? What is generally involved?

Source: From Lighting Controls Association (see Ref. 1).

includes ensuring that (1) light-level or delay-time set points are set, (2) dip switches are set, (3) sensors are aimed for maximum accuracy, (4) preset dimming scenes are set, and (5) the system is tested to make sure it functions as intended. Commissioning and calibration of lighting controls are essential if energy savings are to be achieved and maintained. Occupancy sensors with sensitivity set too high can fail to save energy, but occupancy sensors with too low a sensitivity or too short a delay time can be annoying to occupants. Similarly, improperly adjusted daylight dimming controls can dim the lights too low, causing occupants to override them (e.g., by taping over the sensor), or can fail to dim the lights at all. Choose daylight sensors that can be calibrated quickly and easily, and take the time to calibrate them correctly. The dimming adjustment should be easily accessible to the installer and provide an acceptable range of dimming.

User education: it's often beneficial to educate users about how to use their controls and get them excited.

CONCLUSIONS

Lighting controls are an exciting field in lighting as lighting automation technology rapidly advances and maturity increases reliability and capabilities while decreasing complexity and cost. Lighting automation is proven to save energy while potentially increasing flexibility, security, space marketability, worker satisfaction and aesthetics. A large number of technologies and strategies are available to achieve desired results and performance for virtually any popular application.

RESOURCES

The following web resources can be useful for learning more about lighting and lighting controls:

The Lighting Controls Association:
http://www.aboutlightingcontrols.org

Lightsearch, a directory of manufacturers:
http://www.lightsearch.com

Illuminating Engineering Society of North America:
http://www.iesna.org

Building Codes Assistance Project:
http://www.bcap-energy.org

Advanced Lighting Guidelines:
http://www.newbuildings.org/lighting.htm

Light Right Consortium:
http://www.lightright.org

REFERENCES

1. *LCA Guide to Lighting Control*; The Lighting Controls Association: Rosslyn, VA, 2000.
2. *NEMA Guide to Lighting Controls*; Lighting Controls Association: Rosslyn, VA, 1992.
3. *Typical Lighting Control Applications*; Federal Energy Management Program. http: //www.eere.energy.gov/femp/procurement/eep_light_controls.cfm.

BIBLIOGRAPHY

1. http://www.aboutlightingcontrols.org/education/index.shtml. Series of white papers about lighting controls application and technology.
2. DiLouie, C. *Dimming Ballasts: Technology, Methods, Protocols*. www.AboutLightingControls.org, 2004.
3. Jordan, S. Lighting Control 101. *EC&M*, 2001.

Inv–Light

Lighting Design and Retrofits

Larry Leetzow
World Institute of Lighting and Development Corp./Magnaray International Division, Sarasota, Florida, U.S.A.

Abstract
Lighting designers must consider certain parameters in all retrofit lighting projects. They must also examine specific needs for specific (task) applications. Lighting systems should seek to maximize good lighting quality as well as energy efficiency.

INTRODUCTION

This article will provide overall general guidelines for the selection of appropriate lighting retrofit systems consistent with maximized lighting quality and energy efficiency. These recommended guidelines are not the only reasonable and practical solutions for each and every application situation, but they should serve as a process starting point to determine the most appropriate lighting system for a particular lighting application based on the task dynamics and minimized life-cycle cost. Recommendations in this article have been proven through years of successful application and performance. However, design details for your particular application should usually be determined with the help of an experienced lighting professional. A list of lighting definitions for terms used in this article is provided in Appendix A.

THE MAIN GOAL: QUALITY OF LIGHT

For any lighting retrofit application (interior or exterior), the main goal will be improved lighting quality that will enhance the client's ability to use the lighting system properly. When lighting designs or retrofits are performed properly, energy and maintenance savings will be the advantageous by product, not the main goal. Simply turning out lights saves energy, but obviously, no one can function in the dark.

FACTORS TO BE CONSIDERED

Many factors are involved in selecting a method of reducing lighting energy. Each lighting application has its own intrinsic needs, and those must be considered prior to any change. Just removing lamps from fixtures, removing lamps while adding reflectors, rearranging fixture locations, or changing lamp wattages may reduce energy, but it may not be the most visually efficient way to make a change.

Glare

The most important consideration in lighting design is glare. Glare is the eye's nemesis for seeing. Controlling glare and glare angles can be difficult if multiple functions are occurring at a single visual location or in different directions from a single location. Using efficient linear (fluorescent) sources generally lowers glare levels and allows the eye to function better.

Color Quality

The next important consideration should be the color quality of the scene. Our eyes work best when balanced color sources are used. The eyes contain rods and cones; cones are used for color sensing, and they sense basically in the red (long), green (medium), and blue (shorter) wavelengths. It then makes the most sense to use a source that has balanced color output in these same areas. Presently, the lamp Kelvin temperature that achieves this goal is the 4100 K lamp source, which is available in linear (fluorescent) or point sources (metal halide).

Uniform Light Levels

The third consideration is uniformity of light levels in the work area. This means that "hot" spots and dark spots are reduced to the greatest extent possible. The work area could be a desk, a parking lot, or anything in between. Each task will obviously be designed in a different manner, but the goals should be the same for any visual task.

Understanding the Human Eye

Lighting designers must understand how the human eye works before they can design an appropriate lighting

Keywords: Visual comfort; Visual acuity; Glare; Uniformity; Light quality; Kelvin temperature; Emergency lighting; Task lighting; Unit power density; System efficacy; Photopic; Mesopic; Scotopic.

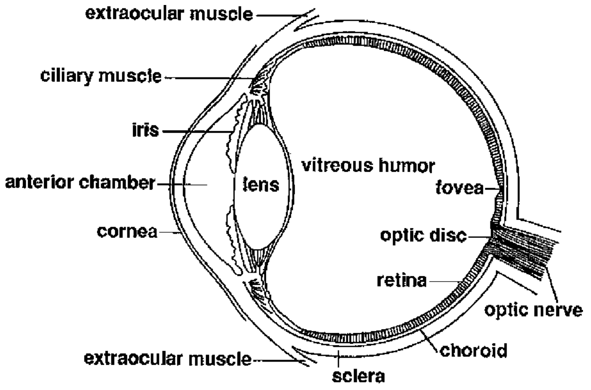


Fig. 1 Shows the architecture of the human eye.

system for a given situation. Understanding the human visual response system and applying this information to outdoor and commercial/industrial indoor designs will allow the designer to provide substantial energy saving opportunities, while also improving visual acuity (Fig. 1).

The central area, called the fovea, contains mainly cones. See Fig. 2.

Rods and cones are spread throughout the entire retinal area. Cones provide color sensing while rods provide acuity, or accuracy sensing. Rods and cones work together in most lighting situations. When both cones and rods are working, the eye is considered to be in mesopia. In extremely low light levels, mainly rods are functioning, and then the eye is considered to be in scotopia. In sunlight, or higher levels of artificial light (the upper level of mesopia has not been determined yet), the eye is said to be in photopia. The eye does not go directly from photopia to scotopia. Mesopia is the transitional area, and mesopic conditions can be experienced in parking lots at night or in many interior spaces during the daytime, including offices, warehouses, and factories.

Figs. 3–5 show how the eye senses different frequencies and levels of light.

Fig. 5 shows the sensitivity ratio. Note how much more sensitive the eye is in the lower light levels to the bluer (whiter) frequencies. Most of today's light meters do not sense this difference, while the human eye does.

(Graphics courtesy GSU Department of Astro Physics).

The scotopic efficacy curve was established in the same manner as the photopic curve. Its sensitivity is shifted, however, to peak at 507 nm, and it decreases in proportionally the same manner as the photopic curve. This results in the scotopic curve reaching a relative value of zero sooner in the visible spectrum than the daylight curve. As a consequence of this fact, our nighttime vision does not see red very well.

The scotopic efficacy curve is assigned a value of unity at 507 nm, and it is represented by the symbol V' lambda. To determine spectral luminous efficacy, the scotopic efficacy value, V' lambda, must be multiplied by 1700 lm (A lumen is the SI unit of luminous flux. Photometrically, it is the luminous flux emitted with a unit solid angle (1 sr) by a point source having a uniform luminous intensity of 1 cd.) per watt (W). This value was adjusted from 1754 to allow both curves to obtain the same value of 683 lm/W at 555 nm. So, a source we see with our dark-adapted vision at 507 nm produces 1700 lm for every watt radiated, and any other wavelength produces a fraction of that value based on the efficacy curve.

If we have a better balance of light colors, we do not need as many lumens (and therefore as many footcandles (A footcandle is a unit of illuminance equal to 1 lm/ft^2 or 10.76 lux.) or lux (A lux is one lumen per square meter ($1 m/m^2$)).) to be able to function visually. If the scene uses dominantly blue light frequencies (mixed with red and green), our eyes can sense details more accurately. However, as mentioned above, glare can cause the reverse, or blinding effects, which make the eye less efficient. Many people dislike the newer (bluer than normal) automobile headlights. Why? Because the eye is more sensitive to blue light at night, and the glare these new lights create is

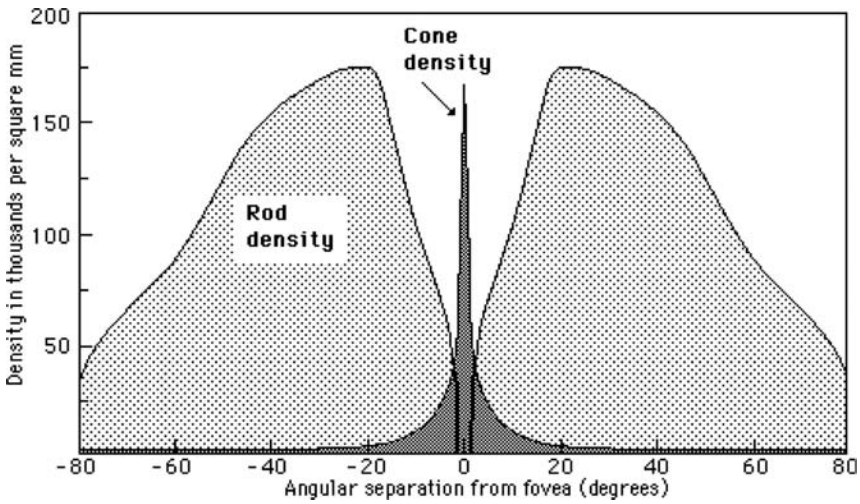


Fig. 2 Retinal rod and cone density chart.

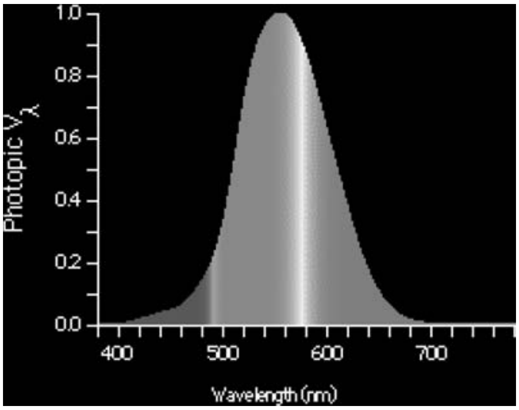


Fig. 3 Photopic (daytime) eye sensitivity curve (CIE). Note peak at 555 nm.

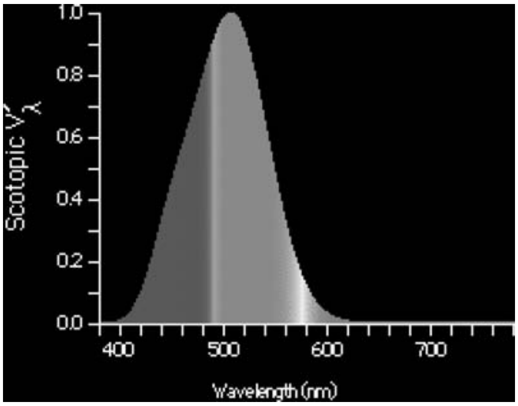


Fig. 4 Scotopic (starlight) CIE. Note peak at 507 nm and smaller amount of the last Section (red in original CIE) in spectrum.

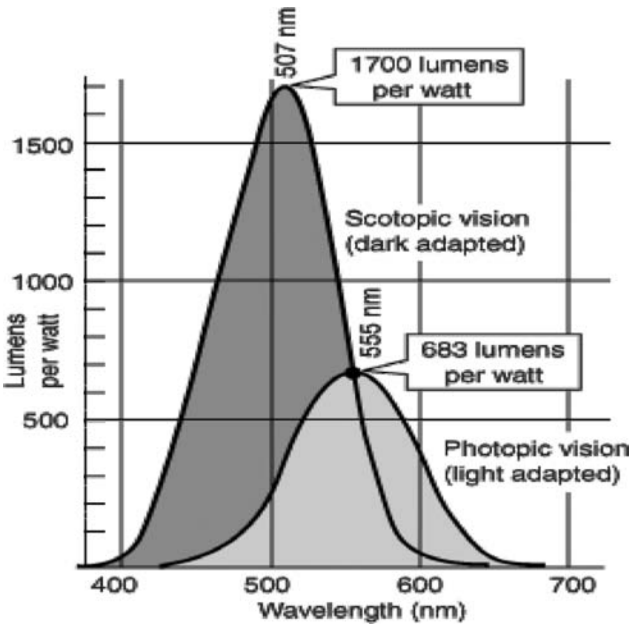


Fig. 5 Scotopic/photopic vision sensitivity chart.

discomforting and sometimes disabling. Standard headlight coloration with "brights-on" also causes discomfort or disabling conditions. Because most cars use the warmer color headlights, these new halogen (cooler color) headlamps stand out more due to their unique coloration. The same factors of color and glare involved in headlights for roadway lighting must be considered for any task, indoors or out.

LIGHT WHERE YOU NEED IT, WHEN YOU NEED IT

There are other factors and systems to consider for conserving maximum amounts of energy, other than light sources and how they are used. The Illuminating Engineering Society of North America's (IES/NA's) motto for this millennium is "Light Where You Need It, When You Need It."

Use of Motion Sensors

Motion sensors add dramatically to energy savings because most people do not turn the lights off when they leave an area. Although motion sensors seem like a good solution for controlling lighting, there are some cases where they are inappropriate. If High Intensity Discharge (HID) sources (high-pressure sodium (HPS), mercury vapor, or metal halide in any form) are used, motion sensors should not be installed due to the long re-strike/warm-up times HIDs require. Motion sensors can be used successfully on fluorescent or incandescent sources. However, incandescent sources are rarely used in new commercial or industrial facilities. If we use motion sensors on fluorescent sources, including all Compact Fluorescent Lamps (CFLs), the number of times the lamp will cycle in a given day must be considered. Fluorescents should not be cycled (turned on/off) more than 12 times/day, or lamp life will suffer significantly. There are a variety of ballast types on the market today that will work properly with motion sensors to provide good longevity for the ballast and lamp. These are called "Program Start" or "Program Rapid Start" ballasts. These ballasts preheat the lamp anodes and cathodes to minimize thermal shock to the lamp filament, thereby increasing the lamp's life substantially. Motion sensors should not be used with screw-in CFLs because their internal ballast is not suited for motion sensor use.

Substantial Lamp Life Increases

Fluorescent lamps using an instant-start ballast have a rating of 14,000 hours. If they use the newer program-start ballasts, they will now have a lamp life of over 24,000 h. These ballasts cost a little more, but are well worth the small extra costs. The newer fluorescent lamps, including twin T5 (PL, Biax, Dulux), linear T5/T5 HO, and T8 lamps, now maintain 90+% in lumen output for 90% lamp life. No other sources maintain that much efficiency to date, including "induction" systems.

Induction Systems

A relatively new lighting system has been developed that uses electrodeless lamp sources with inherent long life results (50,000–100,000 h). The driver or ballast is basically the same type of device that operates a microwave oven to provide light output from the source. The driver/ballast cost is very high at this writing, and it is suggested that the driver be replaced when the source is replaced. These systems should only be considered for extremely difficult replacement applications. At least two manufacturers have sources/drivers product available, and many fixture manufacturers are now bringing products to market. Lumen depreciation should be carefully considered for these systems.

Linear T5 Lamped Systems

Caution must be used when selecting linear T5 lamped units that are enclosed and gasketed. Due to high temperature operation of the linear T5 (not twin T5) lamps, only thermally tested linear T5 fixture systems should be chosen. The small lampholder is the weakest link in the system, and can degrade prematurely. Cracks in the lampholder will affect lamp and ballast life. This can be alleviated or solved by selecting only twin T5 lamped systems, especially for enclosed fixture applications, such as all outdoor applications, and many indoor applications. The larger, more robust lampholder for the twin T5 lamp is better suited for enclosed and gasketed systems.

The case study of Macy's (courtesy of Magnaray® International) in the appendix shows that using motion sensors can provide an additional 25%–30% in energy savings.

Lighting Systems for Dirty Areas

In dirty environments such as factories, warehouses, shop areas, etc. lensed and gasketed luminaires should be used to help maintain higher light levels longer. When luminaries are open, dirt gets onto the lamp and reflector and creates large system inefficiencies by reducing light output. It's like a dirty carburetor on a car—the car still runs, but at reduced efficiency. Cleaning lamps and reflectors is much more expensive than cleaning a lensed and gasketed fixture. Because open luminaires cannot be cleaned while energized, the circuit must be turned off, affecting a large work area where no production can occur. Loss of production time can be very expensive. Lensed luminaires can be cleaned with soap/water (or other cleaning agents) on rolling scaffolding while they are energized, and the cleaning affects only a small area at a time. Production does not have to shut down, and cleaning the lens is much quicker than removing lamps to clean not only the lamps, but the reflectors. The labor savings can be very substantial.

Lensed luminaires keep lamps and reflectors clean longer and maintain higher light levels. The lighting project at Macy's (Brooklyn) mentioned above used enclosed and gasketed luminaries in the shops area. The total savings in this dirty environment was 77% in lighting energy savings. See the case study in the appendix.

Photo courtesy Magnaray® International.

Dimming

Dimming is another somewhat useful tool, but it only saves a portion of the energy that motion sensors save, so it is important to understand why dimming would be better than no light at all. With some of the HID dimming systems, by the time the light is brought from dimmed to full output, the need for increased lighting no longer exists.

PUTTING IT ALL TOGETHER

Good glare control, proper source coloration, and good uniformity are required for a retrofit or a new design to be considered efficient. Once the system is designed, controlling the system's usage can also save substantial energy. An energy management system (EMS) can be used to control lighting with photocells, switches, dimming, or motion sensing.

Each EMS must be programmed properly, and the entire system must be commissioned (checked to make sure it is functioning properly). For verification purposes, periodic reviews of the system operation should be performed, with the timing dependent upon the complexity of the system and the size of the facility. If an EMS is not used, then the proper operation of the components (photocells, motion sensors, switches, etc.) must be periodically verified.

DESIGN PARAMETERS

Each lighting application has a recommended range of lighting levels. In other words, there is no one specific light level that has to be used for any given application. The most recent IES Lighting Handbook/2000 provides a given level of recommended light levels (footcandles or lux) along with a chart or matrix of other factors that must be considered in varying degrees to complete an efficient, quality lighting system design for that application. The final overall design parameters should be determined by a lighting professional to ensure proper benefits and results.

Take parking lot lighting as an example. In various chapters of the IES Lighting Handbook/2000, different levels of light are recommended depending on the various committees' positions. The Security Lighting Committee may call for a higher level of illumination when security is an issue. The parking lot subcommittee of the Roadway Lighting Committee may recommend a low level of light, especially for private parking lots vs public parking lots. The obtrusive light subcommittee may recommend yet a lower light level to minimize, reduce, and restrict light trespass and light pollution. This might lead to the use of controls to have the proper level of light at the time it is needed. Fifty lux (five footcandles) of light might be appropriate while a shopping center is open; but after it closes, these light levels could be reduced by 50% or the lights should all be turned off because no one is on the property. If motion detectors were used with instant-on fluorescent systems, a security alarm could also be connected to the system, and security/police would be notified when the motion sensor activated. High-resolution digital cameras could also be incorporated to record when the sensor activates, and the results could be used for prosecutorial purposes.

CALCULATIONS

Lighting design software is the best way to design a lighting system, but there are other fairly simple ways to estimate or budget for lighting. Methods include wattage per square foot and lumens per square foot. Most people are more familiar with wattages than lumens. We can't see a lumen or a footcandle or a lux in our mind. We do know what a watt is, and that we have to pay for using them. However, both measures have positive and negative aspects.

Let's tackle the negatives first. We have to understand the relationship between watts and lumens. You might think that the more lumens per watt a source possessed, the more light we would get. The best (or maybe worst) example is from a source called low-pressure sodium (LPS). The lamp catalogs say that this source can provide up to 180 lm/W. That sounds good, except for the fact that the eye only uses about 25% of those lumens to send images to our brain, which provide us visual information. If you are a security guard who wants to be able to identify clothing color, hair color, skin color, or the color of a vehicle, you can't do it accurately with an LPS system because all colors appear to be shades of brown, orange, and purple. Therefore, LPS is not a very good source to use for our eye to send accurate images to our brain.

The next most frequently used source today is HPS. Our eyes can only use about 55%–62% of those 90–110 lm/W, but it makes today's light meter read high.

What source is best for our eye? We have already discussed that whiter sources provide better information to our eye and thus more accurate information to our brain so that we can make decisions. This means we should consider metal halide (MH) and fluorescent (FL) sources, whose Color Rendering Index (The CRI is a measure of the degree of color shift objects undergo when illuminated by the light source as compared with those same objects when illuminated by a reference source of comparable color temperature.) (CRI) is much higher than LPS or HPS. Fluorescents lamps have from 60 to 93 CRI, while metal halides are improving, but today still have 50–90 CRIs.

We also discussed that glare needs to be considered in good lighting design. Most of the time, the less glare there is, the better. Point sources (HPS, MH) produce more glare, and should only be considered when FL cannot possible be used. How do we calculate glare? There is no metric today that provides an accurate indication of glare. The Visual Comfort Probability (VCP) doesn't work well for interior applications, and is in the process of revision. The European effort, called Unified Glare Ratio (UGR), which is also under revision, is supposed to work for interior as well as exterior applications. Diffusers, diffractors, and lenses can help reduce glare, but then luminaire efficiency suffers, which reduces system efficiency as well, and energy and operating costs shoot upward.

A new method that compensates for visual efficacy using a multiplier called the Scotopic/Photopic (S/P) ratio is gaining acceptance today. It uses a correction factor to provide a more accurate assessment of what our eye actually senses. A new meter is now finally available (2005) that uses the S/P ratio correction factor and an algorhythm to indicate a visual effectiveness output. The new meter is a radiometer and a photometer combined. The radiometer measures the intensity of each wavelength in the scene (not just the source lamp) and calculates via the S/P ratio the combined amounts of all the energy, then applies that result to the photometer to provide a visual effectiveness measurement. See the chart below for basic multiplying factors for various lamp sources. The visual effectiveness readout can be two to three times higher than just the photopic method of light measurement (today's method of measuring). This means we have been and are wasting from 33 to 66% of our lighting energy costs using today's light meters! This is something we can no longer afford to do. The CIE (Commission Internationale de l'Eclairage), who creates lighting standards, should have developed a *New Method Of Measurement For Mesopic Illumination* in mid—to late 2006. This method of design can actually save the most energy.

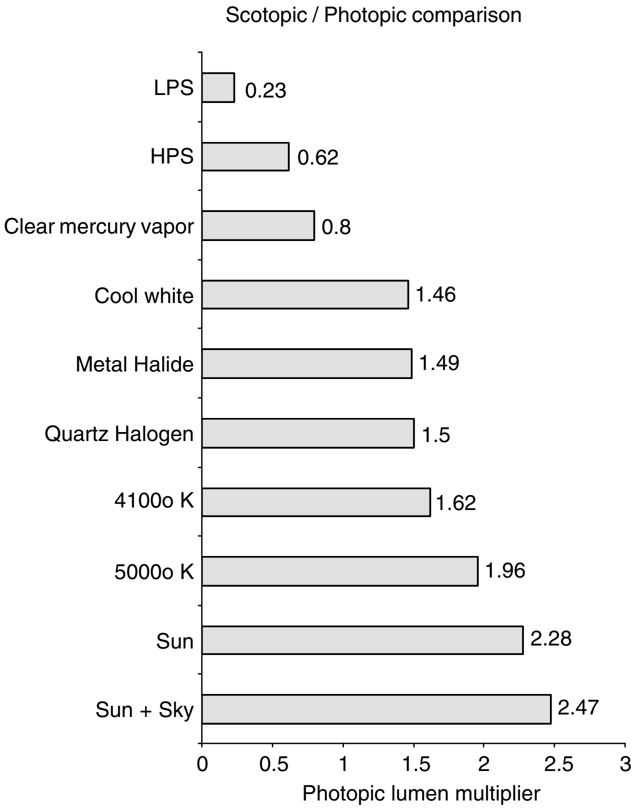


Chart numbers provided by Dr. Sam Berman, Ph.D., FIES.

Emergency Lighting

Emergency lighting must be considered for most applications today, and it is required by many local building codes. When the main lighting system is an HID system, an emergency back up system is required—even for many exterior applications. These systems can be battery/inverter systems, running inverters, backup generators, or even solar-powered back up systems, etc. If the main lighting system is a fluorescent lighting system, backup emergency ballasts can be built into the fixtures. Usually some of the luminaires use emergency ballasts that reduce the light output for a given amount of time. Emergency generators can run larger loads than inverter systems, but normally cost more initially, as well as to maintain. Using the most efficient fluorescent lighting fixtures will reduce generator loads and, therefore, initial costs. One military base changed their main gate entrance lighting system from metal halide lamps (1000 W each) to high-quality, outdoor, weatherproof fluorescent lamps (212 W each), and saved over 40,000 W of generator need. This savings almost paid for the entire lighting system changeover. The author of this article can provide further information about this and other major energy-conserving case studies.

CONCLUSION

To retain the most energy and have the most visually efficient lighting system possible, select the proper lamp source, use the most appropriate luminaire, position it properly for the specific application, and control it with the best device for the specific application. This properly designed system should provide the optimum visual effectiveness level. If handled correctly, a quality designed lighting system will provide tangible and even intangible returns (such as higher productivity) for its owners/users. Proper lighting levels will now be calculated more accurately and it will be easier to provide higher-quality lighting systems, lower maintenance costs, and better value for owners and users. Emergency lighting design can be extremely important for each individual application, and the best methods must be determined for cost and effectiveness.

APPENDIX A

Lighting Definitions

Candela, cd The SI unit of luminous intensity, equal to one lumen per steradian (lm/sr). Formerly *candle*. See Chapter 2 in IES Lighting Handbook (Ref. 1), Measurement of Light and Other Radiant Energy.

Note: the fundamental luminous intensity definition in the SI is the candela. The candela is the luminous intensity in a given direction of a source that emits monochromatic radiation of frequency 540×10^{12} Hz with a radiant intensity in that direction of 1/683 W/sr. The candela so defined is the base unit applicable to photopic quantities, scotopic quantities, and quantities to be defined in the mesopic domain. From 1909 to January 1, 1948, the unit of luminous intensity in the United States, as well as in France and Great Britain, was the international candle, which was maintained by a group of carbon-filament vacuum lamps. For the present unit, as defined above, the internationally accepted term is candela. The difference between the candela and the old international candle is so small that only measurements of high precision are affected. From 1948 to 1979, the unit of luminous intensity was defined in terms of a complete (blackbody) radiator. From this relation, K_m and K'_m, and consequently the lumen, were determined. One candela was defined as the luminous intensity of 1/600,000 m^2 of projected area of a blackbody radiator operating at the temperature of solidification of platinum, at a pressure of 101,325 N/m^2 (N/m^2 = Pa).

Color preference index (of a light source) R_p Measure appraising a light source for enhancing the appearance of an object or objects by making their colors tend toward people's preferences. Judd's "Flattery Index" is an example. See *flattery index*, IES Lighting Handbook (Ref. 1).

Color rendering A general expression for the effect of a light source on the color appearance of objects in conscious or subconscious comparisons with their color appearance under a reference light source.

Color rendering improvement (of a light source) The adjustment of spectral composition to improve color rendering.

Color rendering index (of a light source) (CRI) A measure of the degree of color shift objects undergo when illuminated by the light source as compared with those same objects when illuminated by a reference source of comparable color temperature.

Color temperature of a light source The absolute temperature of a blackbody radiator having a chromaticity equal to that of the light source. See also *correlated color temperature* and *distribution temperature*.

Footcandle, fc A unit of illuminance equal to 1 lm/ft^2 or 10.76 lx.

Footcandle meter See illuminance (lux or footcandle) meter.

Footlambert, fL A lambertian unit of luminance equal to 1/π cd/sq ft. This term is obsolete, and its use is deprecated.

Lumen, lm SI unit of luminous flux. Radiometrically, it is determined from the radiant power as in *luminous flux*. Photometrically, it is the luminous flux emitted within a unit solid angle (1 sr) by a point source having a uniform luminous intensity of 1 cd.

Lux, lx The SI unit of illuminance. One lux is one lumen per square meter (lm/m^2). See the appendix for conversion values.

Mesopic vision Vision with fully adapted eyes at luminance conditions between those of photopic and scotopic vision; that is, between about 3.4 and 0.034 cd/m^2.

Photopic vision Vision mediated essentially or exclusively by the cones. It is generally associated with adaptation to a luminance of at least 3.4 cd/m^2. See *scotopic vision*.

Scotopic vision Vision mediated essentially or exclusively by the rods. It is generally associated with adaptation to a luminance below about 0.034 cd/m^2. See *photopic vision.*

With permission of Illuminating Engineering Society of North America (IES/NA), and is reprinted from ANSI/IESNA RP–16–05—*Nomenclature and Definitions for Illuminating Engineering*

Appendix B

Case Study of Macy's Department Store Shop Area, Brooklyn, NY.

CASE STUDY

Macy's/Brooklyn

Courtesy of Magnaray® International ©Copyright 2000–2006

Macy's Success

Macy's (Brooklyn) basement shops look a whole lot brighter, but they are saving 77% + in energy cost over metal halide originally designed. Felix Tsimmerman of Mader, Smyth, Buyyounouski and Assoc., engineer was very pleased, along with all the folks at Macy's. They loved the use of motion sensors, to take savings from 50% for just lighting load, to the 77% level with sensors.

Everyone thinks it is the best indoor lighting system they have ever seen. While light levels were measured from 50 to 100 "photopic" fc's, the quality and uniformity of the system is *fantasic*. All units were put on motion sensors to save maximum amounts of energy, and reduce maintenance costs.

Inv–Light

Programmed start ballasts keep lamp life where it belongs. A combination of W1PL50, W2PL50, and W4PL50, all having the same multi-lamp, multi-voltage ballast was used. Now only one lamp type and one ballast type are needed in maintenance inventory.

More savings!

Quality is never an accident; it is always the result of high intention, sincere effort, intelligent direction and skillful execution; it represents the wise choice many.

REFERENCES

1. Lighting Handbook, 8th Ed. Illuminating Engineering Society of North America, New York, 1995.

BIBLIOGRAPHY

1. Miller, R.E., II; Tredici, T.J. *Night Vision Manual for the Flight Surgeon*; Armstrong Laboratory: Boston, MA, 1992.
2. Leibrock, C.S.; Reuter, T.; Lamb, D. Molecular basis of dark adaptation in rod photoreceptors, Department of Physiology, University of Cambridge. Eye **1998**, *12*, 511–520.
3. Berman, S.M. Energy efficiency consequences of scotopic sensitivity. J. IES **1992**, *21* (1).
4. Berman, S.M. Spectral determinants of steady-state pupil size with full field of view. J. IES **1992**, *21* (2).
5. Rea, M.S.; Bierman, A.; Bullough, J. Lighting, Mesopic Vision, Spectrum, *Engineering and Applied Sciences*, English 103–385–300 (Last edited on 2004/06/17 12:38:50 GMT-6) 1998.
6. Sagawa, K. 70 Years of Photometry. *Commission Internationale de l'Eclairage*, CIE x-009, 1995.
7. Eloholma, M.; Ketomäki, J.; Orreveteläinen, P.; Halonen, L. Mesopic visibility conditions and luminous environment in driving. In *Proceedings of 25th Session of the CIE* 2003.
8. Akashi, Y.; Rea, M. Peripheral detection while driving under a mesopic light level. In *Lighting Research and Technology*, 1998, Vol. 30.
9. He, Y.; Rea, M.; Bierman, A.; Bullough, J. A system of mesopic photometry. J. Illumin. Eng. Soc. **1997**, *Winter*.
10. He, Y.; Bierman, A.; Rea, M. Evaluating light source efficacy under mesopic conditions using reaction times. J. Illumin. Eng. Soc. **1997**, *Winter*.
11. Bisketzis, N.; Polymeropoulos, G.; Topalis, F. A mesopic vision approach for a better design of road lighting. In *Word Scientific and Engineering Society Transactions on Circuits and Systems*, 2004, Vol. 3.
12. Commission Internationale de l'Eclairage Testing of Supplementary Systems of Photometry, Pub 141, 2001.

Liquefied Natural Gas (LNG)

Charles P. Ivey
Aiken Global Environmental/FT Benning DPW, Fort Benning, Georgia, U.S.A.

Abstract
The world has an abundant supply of natural gas, estimated at roughly 6000 TCF in areas with little or no local demand. This stranded resource has little value unless it can be delivered to areas of high demand that are willing to pay the price of production and delivery. This entry discusses the liquefied natural gas (LNG) export terminals, LNG ships, LNG receiving terminals, and pipelines that make up the chain that can provide a bridge for local communities and industry to access the vast natural gas resources around the world.

INTRODUCTION

What is LNG?

Liquefied Natural Gas is natural gas that has been supercooled to −259°F and changed from a vapor to a liquid. Liquefaction reduces volume by 600 to 1 (Fig. 1). Stored cold in insulated containers at near atmospheric pressure, LNG is colorless, odorless, noncorrosive, and nontoxic. Liquefied Natural Gas becomes lighter than air when vaporized.

HISTORY

Serving as a foundation for this rapidly growing industry is cryogenic technology, which has long-established roots from man's interest in the cold and the supercold. Man's desire to produce ice from water at will apparently is extremely old. The cooling effects of evaporating water, to the point at which ice forms, were known to the ancients of Egypt and India.

During the mid-18th century, scientists were at work in the laboratory on two fronts. Those of an experimental persuasion were engaged in constructing machinery in an effort to produce cold, while those of a more scientific bent were engrossed in trying to derive the gas laws, determine material properties, and unravel the mysteries of work and heat.

Cryogenic technology, thus begun, advanced rapidly under the impetus of two sources: (1) laboratories employing low temperatures for basic studies in chemistry and physics, and (2) suppliers of liquefied gases for industry. The contribution of the suppliers has been the development of new cryogenic processes for separation and purification of gases, and more efficient methods for gas liquefaction. Research laboratories, on the other hand, have been more concerned with the methods of producing the low temperatures needed in their experiments.

LIQUEFACTION

A natural gas liquefaction plant converts the gas received from pipeline or distribution mains into a liquid suitable for transport or storage. The plant consists of two main sections:

1. Feed preparation or pretreatment section—for the removal of carbon dioxide (CO_2), water, and other natural gas constituents that could form solids when the gas is refrigerated and affect the liquefaction process, or adversely affect the quality of the revaporized LNG.
2. Liquefaction section—for the removal of sensible and latent heat from the gas to convert it to a liquid at atmospheric pressure.

Liquefaction of natural gas requires basically the removal of energy in the form of sensible and latent heat. There are several processes available and various modifications of each process that will accomplish this heat removal. The selection of the best process for a particular plant can be made only after a thorough study of the local conditions. In some cases several different processes may fit a particular application.

There are two basic means of accomplishing this heat removal:

1. By transfer through refrigerants to a high-level heat sink, such as cooling water or air. This method uses

Keywords: Liquefied natural gas; LNG; Liquefaction; Vaporization; Regasification; Boiloff; Natural gas; Cryogenic; Storage tank; Tanker.

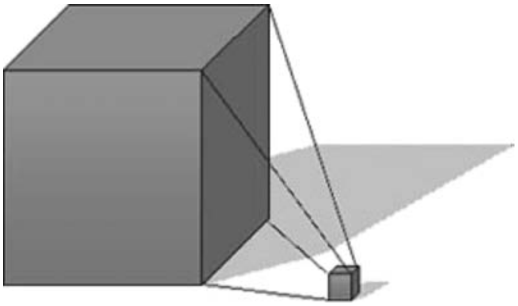


Fig. 1 600-to-1 volume reduction.

one of the following: (a) a cascade cycle using several single-component refrigerants or (b) a modified cascade cycle that circulates a blend of refrigerants in a single refrigerant circuit. Both types use rotating compressors with large fuel gas consumption. The cost of constructing these facilities and operating efficiencies typically determine which process is selected for a given location. Fuel gas can be as high as 8%–12% of the total throughput.

2. By permitting the gas to do work through the use of an expander (Fig. 2). The expander cycle can utilize the pipeline pressure and expand natural gas, or it can use a compressor–expander combination using a separate gas (i.e., nitrogen) as a circulating refrigerant. The expander typically uses less fuel gas when a letdown pressure is used as the energy source. In the past few years, both types of plants have been built and are in operation.

STORAGE

Large volumes of natural gas can be stored in the liquid state because, as a liquid, LNG will occupy approximately 1/600th of its gaseous volume of natural gas under standard conditions. To maintain natural gas as a liquid at atmospheric pressure, it is necessary to maintain the temperature at approximately −260°F (−162°C).

Because refrigeration at such temperatures is costly, most storage systems rely on very well-insulated, nonrefrigerated containers. For large-volume storage, the internal tank pressure is usually limited to near atmospheric for structural and economic reasons. Storage of smaller volumes of LNG at various levels of positive pressure can be achieved, typically using vacuum-jacketed storage tanks.

With any type of container used to store LNG, there will be a certain inflow of heat to the container, resulting in evaporation of the product. For typical LNG storage tanks, this vaporization or boiloff rate will range from about 0.04% to 0.2% per day of the total storage volume,

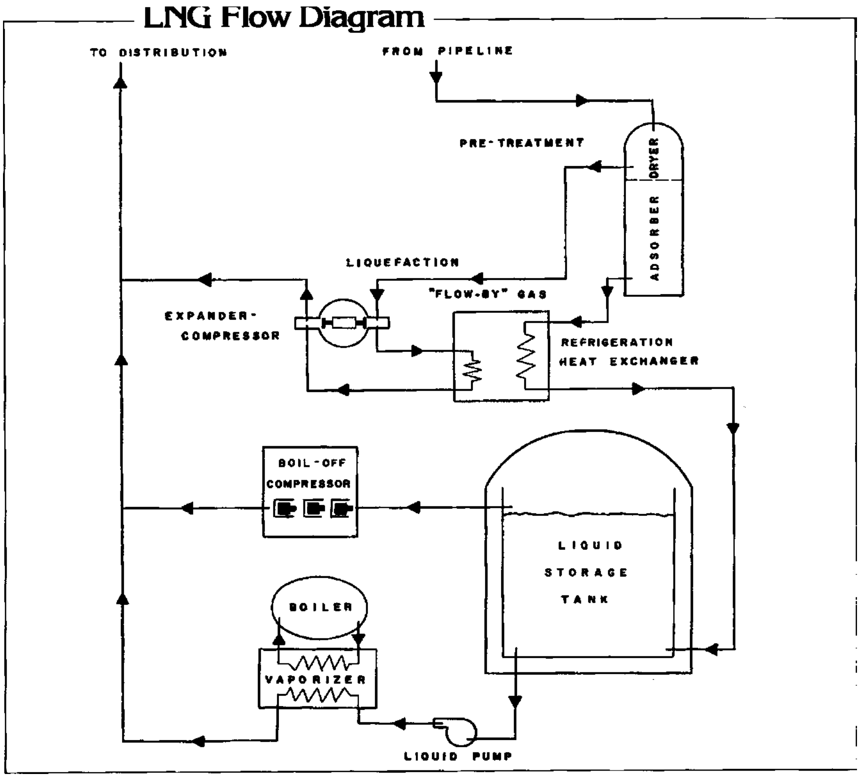


Fig. 2 Simple flow-by expander–compressor liquefaction cycle diagram.

Liq–Mobil

Fig. 3 Liquefied natural gas (LNG) storage tank with remote heat vaporizer in right foreground. Note relief valve stacks on tank top.

depending on the type of storage, the amount of insulation used, and the size of the container. In addition to boiloff due to storage heat gain, flash vapors can be generated in the storage tank during liquefaction, resulting from pressure reduction on LNG that is not subcooled before entering the tank from the liquefaction plant. Both occurrences generate vapors that need to be recovered using compressors to prevent atmospheric venting.

Two areas of interest have indicated a definite need for large-volume storage of LNG. First, in transporting gas from energy-rich to energy-poor areas of the world in insulated tankers, the need for large storage depots at both the loading and the market ports is evident. In this case, it is not necessary to preserve the LNG in the tank for long periods, because it is passed on to distribution within a short period. Consequently, higher heat influx resulting in greater boiloff rates can be tolerated, as the vaporized LNG can be used as part of the sendout gas or as fuel.

In the second area, the continued growth and expansion of the natural gas industry have created the necessity for large-volume storage near metropolitan areas to meet winter peak loads that exceed pipeline capacity, also known as peak-shaving. In this case, LNG is stored for relatively long periods and used (i.e., vaporized and distributed) during only a few days of the winter. Consequently, heat influx must be held to a minimum.

Whether the LNG plant is to be used for base-load operation or peak-shaving, a large portion of the cost is the storage facility (Fig. 3). Also, the large volume of storage associated with LNG plants represents a high concentration of energy. It is for these reasons (i.e., economics and reliability) that considerable research and development expenditures have been made relative to storage concepts.

Conventional above-ground, double-wall metal tanks have been used successfully for cryogenic service since the early 1950s. These tanks have generally been flat-bottomed, cylindrical, dome-roof, double-wall tanks with an outer shell of carbon steel or concrete and an inner tank of 9% nickel steel or aluminum that are capable of retaining their strength properties at these cryogenic temperatures. Below −40°F, the loss of strength properties of carbon steel eliminates it as a suitable LNG containment material. Insulation between the two shells usually consists of perlite in a nitrogen or natural gas atmosphere.

The double-wall, flat-bottomed tank—the conventional configuration for above-ground metal tanks—is really a tank within a tank, with the annular space between the two filled with insulating materials of various types. The inner tank, in contact with the LNG, is made of materials suitable for cryogenic temperatures of −260°F (−162°C) and the design loadings of the LNG. The outer tank serves the primary purpose of containing gas pressure and insulating materials that surround the inner tank. The outer tank provides a proven means of protection for the insulation system against external forces, such as fire, impact, and—most important—weather and moisture. The outer tank is vapor-tight, so as to contain the desired internal gas pressure and preclude migration of moisture from the atmosphere into the insulation materials. Connections to the inner tank must pass through the outer tank and are of a special design to permit differential movements due to varying temperature and hydrostatic load.

Liquefied Natural Gas tank foundations must be more carefully designed and constructed than foundations for chemical and petroleum tanks. Limits on construction tolerances and foundation settlements are much more restrictive and important for three reasons:

1. Liquefied Natural Gas tanks are 2–4 times as high (60–120 ft or 18–36 m) as petroleum tanks (30–60 ft or 9–18 m). Any differential foundation settlements can cause the tank to display out-of-roundness and possible uneven liquid levels relative to the tank shells. These conditions are worsened as tank shell heights increase.
2. Internal pressure of LNG tanks requires that the tanks be anchored to the foundations to resist uplift. For anchorage and support of a tank to work properly, the loads must be carried uniformly by the foundations.
3. Load-bearing insulation must be supported uniformly on a nearly flat plane to prevent cracks from forming in the insulation with consequential high heat leak into the LNG.

If the tank foundation is of the ring-wall type, the outer bottom will rest on a gravel or sand base directly upon the soil. Because there is a certain amount of heat migration through the bottom, tank heating coils are installed under the bottom to prevent any soil freezing and consequent frost heaving. In the case of a foundation supported on piles, a minimum 18-in. (460 mm) air space is provided between the ground surface and the bottom of the pile cap to allow for the circulation of air, which eliminates the need for any additional heat source.

In LNG and cryogenic storage systems, insulation is a primary factor. Proper insulation is the key to successful and economical storage. The outer steel shell of the double-wall LNG tank is the best possible vapor barrier against migration of moisture. It permits a gas-tight insulation space where a dry atmosphere can be maintained. This dry atmosphere in a suspended-deck tank is provided by the boiloff from the inner tank.

The insulation material used in double-wall LNG tanks is an inert, inorganic, noncombustible, granular material commonly called perlite. The installed thickness of the perlite governs the overall rate of heat influx through the system. The amount of heat influx into the tank determines the factor commonly called boiloff, which refers to the percentage of the tank's contents that will evaporate in a given period under specified ambient temperature conditions. The term normal evaporate rate (NER) is sometimes used.

The problem of insulating the space between the tank bottoms is different from that for the annular space between the walls and roof. In the region between the tank bottoms, a load-bearing type of insulation is used. Cellular glass block has proved to be satisfactory for this service, but care must be taken to avoid damaging the somewhat-fragile material during installation and during subsequent periods when tank erection work is under way. Care must also be taken to prevent excessive moisture from penetrating the blocks during construction.

REGASIFICATION/VAPORIZATION

Regasification and sendout of LNG are the final steps in the operation of an LNG peak-shaving plant or import terminal. The regasification or vaporization is accomplished by the transfer of heat to the LNG from ambient air or water, fuel combustion in integral or remote vaporizers (Fig. 4), or recovered waste or process heat. The regasification system, including cryogenic pumps for LNG, must be highly reliable and capable of operating continuously for long periods as well as for brief periods of peak demand.

The cost of the regasification system, also known as the vaporization system, generally represents only a small fraction of the cost of the entire storage plant; however, reliability of the system is most important, because failure or breakdown could result in disruption of service at a crucial time.

The regasification section of a peak-shaving plant may be designed for only a few days of operation during the winter to meet extreme peak loads. To obtain adequate reliability, total sendout capacity may be divided into several independent parallel systems, each capable of handling all or a substantial fraction of the total demand.

Because of the large quantities of liquid and gas being handled and the need for extreme reliability, regasification systems must be engineered carefully. Liquid natural gas, as well as cold vapor, must be prevented from passing into the distribution system. Hydrostatic surge loads must be considered in the liquid piping system. Therefore,

Fig. 4 Remote boilers used in regasification process.

instrumentation for process measurement, control, and safety are important considerations.

Regasification requires the following major operations:

1. Pumping LNG from storage to distribution or transmission system pressures.
2. Vaporizing liquid to gas.
3. Controlling process flow, pressure, and temperature.
4. Odorizing and metering the sendout stream.

The physical surroundings, sendout rate, and specific type of LNG facility (base-load as compared with peak-shaving) would determine which method of vaporization is most economical. A base-load plant, which normally would run continuously, can justify a higher investment cost to achieve lower operating cost. On the other hand, with a peak-shaving plant, which would operate for only brief periods during the year, a much lower investment cost would be possible because a higher operating cost is allowable.

The different vaporizers available can be divided into four general categories: (1) integral-heated (fired), (2) remote-heated (fired), (3) ambient-heated, and (4) process-heated.

Although the design and economic considerations discussed above are important in the selection of equipment for base-load operation and peak shaving, the requirement for base-load continuous service is perhaps the most critical. Care must be exercised in evaluating vaporizer and pump performance in peak-shaving plants, where the service period ranges from days to weeks per year. Base-load LNG plant equipment must have service reliability in excess of 8000 h per year. This requires that special consideration be given to such items as bearing life in pumps and selection of materials in hot or corrosive areas, such as combustion zones or those exposed to products of combustion.

There is increased interest in floating and other offshore LNG liquefaction plants, and proposals have been developed for floating or offshore LNG import terminals as a solution to siting problems in the United States. Requirements for base-load operation offshore are similar to those for onshore operation, although compactness, safety, reliability, and low maintenance may be even more critical. Because power generation is inherently required for offshore facilities, the recovery of gas turbine or power generation waste heat is available for supplemental vaporization, along with the primary seawater, although the trade-off of complexity vs efficiency must be analyzed.

If all the low-temperature source available at a base-load receiving terminal could be used effectively over the entire temperature range from −260°F (−162°C), the value of the refrigeration would approach the initial cost of liquefaction. It is not practical to recover all the cold, but a large number of schemes have been proposed or installed to take partial advantage of the refrigeration available, which leads to reduced vaporization cost.

An example of a cryogenic recovery or cold utilization vaporizer system is one that would appeal to total-energy companies or utilities that supply both natural gas and electrical power. The gas source would be LNG, and the power would be generated with gas-fired turbines.

The process can be described as having the following features:

1. The production of horsepower by means of a gas-fired turbine driver.
2. The continuous vaporization of LNG with the large quantity of low-quality heat available from the gas-fired turbine exhaust.
3. The use of LNG or low-temperature natural gas for the purpose of precooling the inlet combustion air to the turbine, thereby increasing the power output of the generator.

The Eco-Electrica LNG plant in Puerto Rico has installed such a system for the purpose of generating electrical power for offsite consumption. The idea has been widely studied and appears to have merit.

Although the sendout or regasification section of an LNG storage facility is one of the least costly plant elements, it is probably the most important, because if it should fail to operate on the few occasions when it is called upon for peak demand, the entire purpose will have been defeated. When peak demands occur, the regasification system must be ready to operate immediately and reliably. Fuel gas for regasification typically uses about 1.5%–2% of the facility throughput of natural gas.

In a base-load plant, which is designed to operate a full 365 days per year, interruption of the regasification section has equally serious consequences. A base-load import terminal is just one critical link in the import chain from gas production and liquefaction to shipping and terminal operation. Failure of any link leads to an interruption of gas supplies and may involve high costs of "take or pay" contract provisions to protect the investment in such capital-intensive projects from loss of income during periods of unscheduled downtime.

TRANSPORTATION

All LNG is transported in one manner or another, whether a distance of a few hundred feets by pipeline from liquefaction to storage or several thousand miles between countries by ocean tankers. For intermediate distances, specially designed pipelines are used for loading and unloading LNG tankers, and have been proposed for transporting LNG greater distances overland. Liquefied Natural Gas has been transported by tanker truck in the United States for several years. For transportation of larger quantities of LNG, barges designed for use on inland waterways and for use on the open seas have been proposed. All these modes of transportation can be classified into two broad categories: marine transportation and overland transportation.

Marine transport of LNG has advanced considerably since the first LNG cargo ship technology was developed in the late 1950s and early 1960s. Since that time, a large number of LNG containment systems has been studied and developed. The early designs that were commercialized have proved to be safe, reliable systems for marine transport of LNG. In addition, new designs and system modifications are under development and greatly benefit from the use of vacuum-jacketed LNG storage tanks in lieu of external insulation.

Liquefied Natural Gas ship (Fig. 5) size is spoken of in terms of cubic meter of total liquid capacity. Therefore, ship size will be given in cubic meter of liquid only (1 m^3 equals 35.31 ft^3 or 6.29 bbl liquid). Liquefied Natural Gas tankers range in size from about 5000 m^3 for the early pioneering designs to the 125,000–145,000 m^3 ships that have become the industry standard. Ships of 165,000 m^3 have been designed, and a 330,000 m^3 ship has been proposed. Most ships are constructed in Korea at this time.

Typically, LNG tankers are loaded directly by cryogenic pumps from the onshore storage tanks. During loading, excess vapor is generated within the cargo tanks, while the dropping liquid level in the storage tank results in a need for vapor to prevent creation of a vacuum; therefore, excess vapor from the cargo tanks is returned to the storage tank via an insulated vapor return line. Excess vapor may be absorbed by the system, consumed for fuel, or vented to the atmosphere. Unloading is accomplished in a similar manner through the use of submerged cargo pumps within the cargo tanks.

Liquefied Natural Gas can be transported overland in pipelines or mobile tanks. Historically, short-distance, large-volume transport has been through insulated pipelines, as from liquefaction plant to storage or in loading and unloading tankers. Smaller volumes can be moved over longer distances in insulated over-the-road trucks or railway cars. Periodically, long-distance LNG pipelines have been proposed but have been rejected due to the high cost of sufficiently efficient insulation systems.

Prior to 1969, only a few small-scale over-the-highway movements of LNG had been attempted in this country using equipment originally designed for liquid nitrogen service. Based on the success of these initial operations and the increasing need for highway transport of LNG, new equipment designed and built specifically for LNG service became commercially available for the first time in the early 1970s. This equipment is produced by companies that have long provided trailers for the industrial gas industry. The design of this new equipment was based on well-matured cryogenic technology, and incorporated reliability and safety concepts developed through years of experience in transporting liquid nitrogen and oxygen.

Fig. 5 Liquefied natural gas (LNG) transport ship.

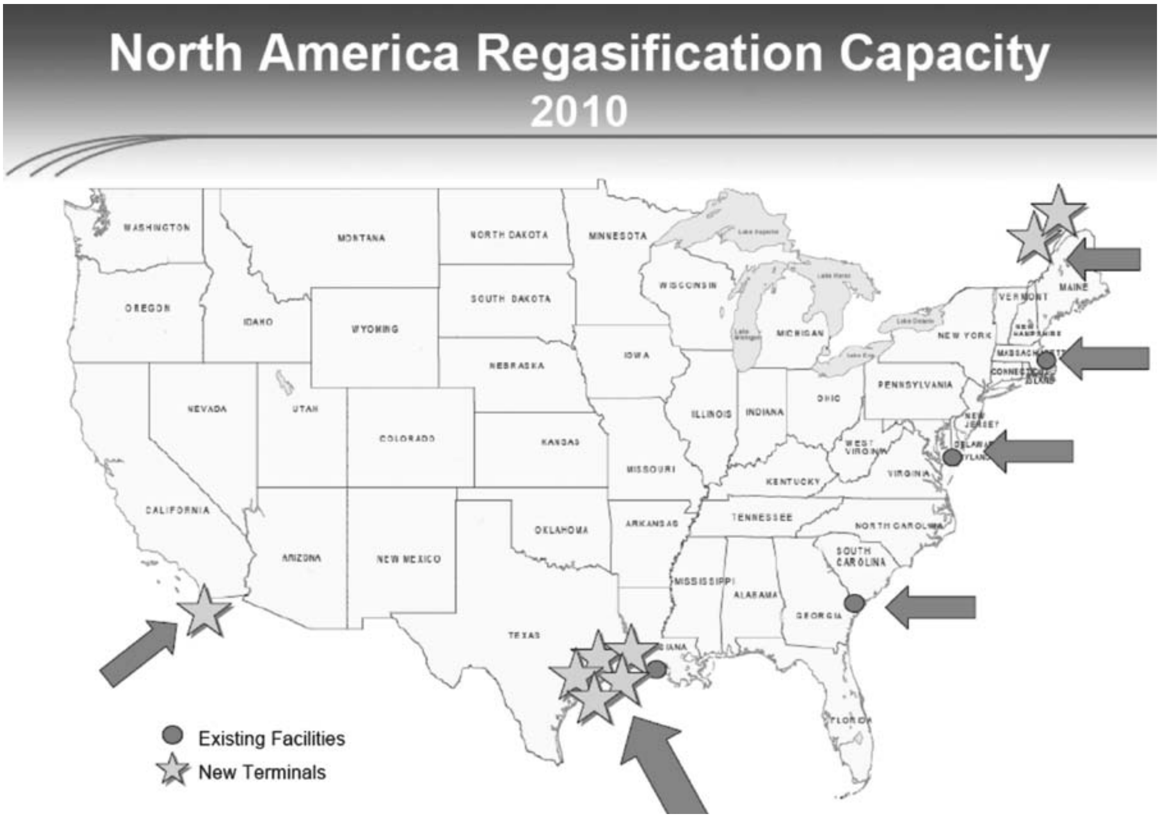


Fig. 6 These new facilities will play a vital role in meeting existing and projected natural gas demands.

In the ensuing years, the widespread highway movement of LNG has become commonplace and has established an enviable record of safety and reliability.

CONCLUSION

The LNG industry has been a key player in the U.S. natural gas supply system since the early 1960s, used primarily in a peak-shaving role. It has been proved that the liquefaction, transportation, storage, and regasification of stranded resources of natural gas constitute an economically feasible solution to declining domestic production and resulting price volatility due to constrained supplies and increasing demand—all this with an enviable safety record.

Based on predictions by Cheniere Energy (www.cheniere.com), the next phase in the maturing of the U.S. Liquefied Natural Gas industry will be in the base-load import terminal area. Currently, 17 liquefaction export plants and more than 150 LNG marine tankers are operating around the world. With more than 25 new liquefaction export terminals proposed to be built prior to 2010 and more than 100 marine vessels slated for delivery before 2008, the stability of the LNG commodity trading market is becoming more secure with each passing day.

The current bottleneck in the U.S. market is receiving capacity. Fig. 6 shows existing and proposed new U.S. import terminals with projected completion dates prior to 2010. Public education and acceptance play a key role in market development.

> "Our limited capacity to import liquefied natural gas effectively restricts our access to the world's abundant supplies of natural gas." (Alan Greenspan, U.S. Federal Reserve Bank chairman, May 21, 2003)

> "What we need to do is get in place, as soon as we can, the capability of fairly substantial imports that enable our manufacturers who use natural gas to compete internationally." (Alan Greenspan, U.S. Federal Reserve Bank chairman, April 21, 2004)

BIBLIOGRAPHY

1. A brief history of U.S. LNG incidents, www.ch-iv.com/lng/incid1.htm.
2. www.ENERGY.CA.GOV/lng.
3. www.lngfacts.org/
4. www.ferc.gov/industries/lng.asp.
5. www.ferc.gov/for-citizens/lng.asp.
6. www.naturalgas.org/lng/lng.asp.
7. www.eia.doc.gov/pubs/oil_gas/natural_gas/features_articles/2003/lng/lng2003.pdf.

Living Standards and Culture: Energy Impact

Marc A. Rosen
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Abstract
Energy is linked to living standards and culture in complex ways and these relations are examined in this article. First, living standards and culture are described. Energy use is described and its relation to population and urbanization is examined. The environmental impact of energy use is described and the impact of energy use on society, living standards, culture, and sustainability is discussed. Finally, possible energy-use modifications to improve living standards are addressed, including increased efficiency, fuel and energy-resource substitution, and strategic planning. The differences between developing and developed countries are addressed throughout.

INTRODUCTION

Living standards and culture are related to energy even though the relations are often complex and sometimes difficult to discern.

Living standards loosely defines the degree of material comfort available to a community. The availability of energy resources and the ability to utilize them strongly influence a society's living standards.[1–6] Living standards also have a feedback effect on energy issues. For example, societies with high living standards likely have good education systems and extensive research and development undertakings that permit the development of energy technologies capable of harnessing energy resources more efficiently and with less environmental impact.

Culture is usually loosely thought of as the form and stage of intellectual development or civilization. Energy choices are sometimes dependent on a society's culture, while at other times, energy-related factors contribute to cultural changes and development.[4,5]

Environmental impact is often a significant consequence of energy use[7–18] and it strongly affects and is affected by living standards and culture. Environmental issues also affect the sustainability of a country's development in the longer term[19–22] and thus are an important consideration in discussions of living standards and culture. This is seen in historical data for global energy-resource use (Table 1) and environmental data and related environmental emissions (Table 2). For example, increasing use of fossil fuels is seen to correlate with increasing CO_2 concentration in air.

In this article, the impact of energy on living standards and culture is examined. First, living standards and culture are described. Then, energy use—in the world and in countries—is described along with its relation to population and urbanization. The environmental impact of energy use is described. Then, the impact of energy use on society, culture, living standards, and sustainability is discussed, and possible energy-use modifications to improve living standards are covered.

LIVING STANDARDS AND CULTURE

A living standard is defined as the "degree of material comfort available to person or class or community" and as "a level of subsistence, as of a nation, social class, or person, with reference to the adequacy of necessities and comforts in daily life." The social and economic indicators that contribute to living standards have been examined,[1–6] and some of these indicators are listed for selected countries in Table 3 and Table 4, respectively.

Culture is defined as the "particular form, stage, or type of intellectual development or civilization" and as "the concepts, habits, skills, art, instruments, institutions, etc. of a given people in a given period," or just as "civilization." Factors that contribute to culture include its standards for greetings and dress, social taboos, customs and traditions, crafts, local foods, and architecture.

Living standards and culture can affect each other. For example, a society with high living standards may have ample free time to devote to cultural development and consequently may develop a wide range of arts and skills while a society with lower living standards may focus on the development of practical skills associated with basic necessities. Also, the cultural choices made by a society (e.g., valuing wealth) can affect living standards.

Keywords: Living standards; Culture; Energy; Environment; Pollution; Sustainable development; Energy conservation.

Table 1 Selected global energy-resource utilization data

	Year							
	1860	**1880**	**1900**	**1920**	**1940**	**1960**	**1980**	**2000**
Fossil fuel use (millions of tons of oil equivalent per day)[a]								
Oil						800	2,800	3,400
Coal						1,500	1,800	2,500
All fossil fuels						2,400	6,000	8,000
Primary energy supply (millions of barrels of oil equivalent per day)	0.3	3	13	18	20	42	110	170

[a] Data are only available in source for last half century, and data for 2000 are based on extrapolations of data for the 1990s.
Source: Adapted from Island Press and Prentice Hall (see Refs. 1,2).

ENERGY USE

Population and Energy Use

World population is expected to increase from 6.2 billion in 2003 to about 10.5 billion in 2050.[23–26] Economic development will likely continue to grow, with global demand for energy services expected to increase from 1990 levels by as much as an order of magnitude by 2050, while primary-energy demands are expected to increase by 1.5–3 times.

The world population's share of developing countries is about 77%, and this is expected to reach about 85% by 2050. Yet developing countries are responsible for only a quarter of global energy use. The ability to provide energy services for the developing world must grow considerably to meet the extra demands expected in these countries and to ensure that their economic development is not constrained.

The energy needs of cities are large and increase with both urban growth and industrial development. In general, urbanization entails not only major changes in land-use patterns, but it also shifts in the ways societies use energy. The transition away from traditional fuels (e.g., wood) to fossil fuels and electricity is accelerated with urbanization. Urbanization is accelerating in many developing countries.

Energy-Use Growth Patterns

Several methods are used by the International Energy Agency to project future energy use.[23] Using a "capacity constraints case" (Table 5), world demand for primary energy increased by more than 44% between 1992 and 2010—corresponding to an average annual rate of about 2.1%—to 11,489 million tonnes of oil equivalent (Mtoe). The average annual growth rate through this period for natural gas is 2.5%, the fastest among all fossil fuels, and

Table 2 Selected energy-related emissions and environmental data

	Year							
	1860	**1880**	**1900**	**1920**	**1940**	**1960**	**1980**	**2000**
Energy-related emissions								
Carbon emissions from burning fossil fuels (million tons)[a]								
Industrial countries						1,500	2,600	3,000
Developing countries						470	1,100	2,800
Former East Bloc countries						500	1,200	700
World						2,400	5,000	6,600
Environmental data								
Average temperature at earth's surface (°C)	14.5	14.6	14.8	14.7	15.0	15.0	15.1	15.3
Atmospheric CO_2 concentration (parts per million)	283	290	292	298	308	311	330	370

[a] Data are only available in source for last half century, and data for 2000 are based on extrapolations of data for the 1990s.
Source: Adapted from Island Press and Prentice Hall (see Refs. 1,2).

Table 3 Key social indicators for selected latin and caribbean countries (1998)

	Argentina	Bolivia	Brazil	Colombia	Domincan Republic	Haiti	Honduras	Mexico	Nicaragua
Population									
Total population (millions)	36.1	7.9	165.9	40.8	8.3	7.6	6.2	95.8	4.8
Average annual growth (1992–1998)	1.3	2.4	1.4	1.9	1.8	2.1	2.9	1.7	2.8
Urban population									
% of total population	89	61	80	73	64	34	51	74	55
Annual growth	1.6	3.3	2.0	2.5	2.8	3.9	4.9	1.9	3.2
Life expectancy at birth (years)	73	62	67	70	71	54	69	72	68
Infant mortality (per 1000 live births)	19	60	33	23	40	71	36	30	36
Child malnutrition (% of children under 5)	2	8	6	8	6	28	25	n.a.	12
Access to safe water									
Urban[a] (% of population)	71	78	85	88	74	37	81	91	81
Rural[a] (% of population)	24	22	31	48	67	23	53	62	27
Poverty (% of population below poverty line)	18	n.a.	n.a.	18	21	n.a.	53	n.a.	50
Illiteracy (% of population over age 14)	3	16	16	9	17	52	27	9	32

n.a. = not available.
[a] Year of information varies (1991–1995).
Source: Adapted from The World Bank Group (see Ref. 21).

Table 4 Key economic indicators for selected latin and caribbean countries (1998)

	Argentina	Bolivia	Brazil	Colombia	Domincan Republic	Haiti	Honduras	Mexico	Nicaragua
GNP/capita (Atlas method, U.S.$)	8030	1010	4630	2470	1770	410	740	3840	370
GNP (Atlas method, U.S.$)	290.3	8.0	767.6	100.7	14.6	3.2	4.6	368.1	1.8
Industry contribution to GNP									
% of GNP	28.7	28.7	28.8	25.1	32.8	20.1	30.9	26.6	21.5
Average annual growth	3.2	n.a.	0.5	−2.3	8.8	6.1	9.0	6.6	4.6
Services contribution to GNP									
% of GNP	65.6	55.9	62.8	61.4	55.6	49.6	48.8	68.4	44.4
Average annual growth	4.7	n.a.	1.3	2.0	7.7	2.4	5.9	4.5	3.6

n.a. = not available.
Source: Adapted from The World Bank Group (see Ref. 21).

Table 5 Past and projected global energy consumption (based on IEA capacity-constraints case)

	Year			
	1971	**1992**	**2000**	**2010**
Primary energy (Mtoe)				
Solids	1,510	2,301	2,612	3,280
Oil	2,327	3,109	3,549	4,394
Gas	896	1,745	1,979	2,708
Nuclear	29	554	658	705
Hydro	104	192	245	312
Geothermal/others/renewables	4	34	57	90
Total	4,870	7,935	9,100	11,489
Final energy (Mtoe)				
Solids	814	915	1,040	1,255
Oil	1,899	2,602	3,019	1,780
Gas	582	1,002	1,128	1,438
Electricity	461	1,048	1,276	1,747
Total	1,756	5,567	6,463	8,220
Transformation and losses (Mtoe)	1,114	2,368	2,637	3,269
Electricity generation by source (TWh)				
Solids	2,165	4,774	5,946	7,991
Oil	1,100	1,387	1,313	1,406
Gas	717	1,652	2,358	4,423
Nuclear	111	2,126	2,520	2,707
Hydro	1,210	2,235	2,846	3,630
Geothermal/others/renewables	5	46	83	166
Total	5,308	12,220	15,066	20,323

Source: Adapted from Organization for Economic Cooperation and Development and International Energy Agency (see Ref. 23).

world consumption of coal and other solid fuels increased annually at an average rate of 1.6% to 3067 Mtoe. Hydraulic and nuclear energy are also expected to remain important, and the use of renewable energy sources is expected to reach about 100 Mtoe by 2010. Selected global environmental and energy-related socioeconomic data for the "capacity constraints case" is presented in Table 6.

Table 6 Past and projected global energy-related socioeconomic and environmental data (based on IEA capacity-constraints case)

	Year			
	1971	**1992**	**2000**	**2010**
Energy-related socioeconomic data				
GDP per capita (1987 U.S.$)		3,511	3,938	4,699
Energy use per capita (toe)	1.30	1.46	1.48	1.64
Energy intensity (toe/1000$)		0.43	0.39	0.36
Environmental data				
CO_2 emissions (Mt)	14,707	21,114	24,073	30,726
CO_2 emissions (% change since 1990)		−2.4	11.3	42.1

Source: Adapted from Organization for Economic Cooperation and Development and International Energy Agency (see Ref. 23).

ENVIRONMENTAL IMPACT OF ENERGY USE

Environmental concerns associated with energy use impact living standards and range from pollutant emissions, hazards, and accidents to the degradation of environmental quality and natural ecosystems.[7–18]

In the 1970s, concerns about energy use mainly focused on economics and the availability of a reliable supply of energy resources. Most countries began to address environmental problems in the 1980s, adopting laws and policies aimed at coordinating economic development with environmental concerns. Concerns over energy-related environmental issues have come to be part of the culture in some countries.

Energy-related environmental concerns have expanded from being primarily local or regional to being multinational and global, with some of the most significant problems like global climate change and ozone depletion falling into the latter category. In developing countries and in countries with emerging industrial economies in particular, growth rates in energy consumption are typically high and extensive environmental management practices are not yet fully adopted. At present, however, industrialized countries are responsible for most air pollution, ozone depletion, and carbon emissions. Contributions from developing countries are smaller but likely to increase as they further industrialize.

Decisions regarding energy policy alternatives require comprehensive environmental analysis. Problems such as acid precipitation can be dealt with in part by technical and regulatory measures. For example, societies can implement vehicle exhaust standards or emission limits for power stations. Such measures impact a relatively focused and small number of parties. However, sources of greenhouse gas emissions are widespread and the gases readily disperse over large geographic areas. Thus, local and relatively limited approaches to dealing with global climate change are not generally effective and comprehensive energy policies are needed that influence energy consumers and producers in all countries.

Increasing energy efficiency can somewhat slow growth in energy use and carbon emissions, although it likely cannot offset the increasing energy use in many developing countries to support economic growth. Significant decreases in the energy requirements for economic development probably require more fundamental changes in such societies and the way they develop. Industrialized countries, for example, usually exhibit continuously increasing energy use and carbon emissions due to the desire for greater comfort and convenience. Many developing countries have followed the development of industrial ones, often leading to urban blight and other problems.

Depending on the culture and the values of a society, energy measures can sometimes be introduced voluntarily and have substantial success. In other situations, governments must use incentives and enforcement measures such as laws and penalties to achieve significant benefits. For example, the combination of inexpensive energy supplies and not too stringent environmental constraints in North America has led to a culture of travel by automobile and less of a preference for public transit. To substitute public transit for automobile use in that society is therefore difficult.

IMPACT OF ENERGY USE ON LIVING STANDARDS, CULTURE, AND SUSTAINABILITY

Energy clearly impacts living standards and culture, while these topics in turn often affect energy choices. Natural energy—direct solar radiation, its derivatives such as wind and wave energy, as well as geothermal and tidal energy—makes possible the existence of life, ecosystems, and human civilizations. Additional energy, which includes the secondary flows of energy produced by humankind, contributes to advanced technological stages of production and influences the evolution of living standards. Civilizations generally adapt to their environments, developing their own systems of values, consumption patterns for energy and other resources, and development paths.

An abundance of energy resources can help a society achieve high living standards and economic prosperity simply through harvesting the resources, although energy-related environmental degradation can also result. By extension, cultural choices, directions, and development each can be affected by the availability of energy resources. Yet, the possession of abundant energy resources does not always lead to high living standards and countries that have little or no domestic energy resources can often achieve high living standards, often through developing a culture that highly values learning, knowledge, and innovation.

Significant disparities exist with wealth and living standards between developed and developing countries. For example, statistics of modern life standards show that per capita incomes of the populations of some less-developed countries are less than 1% of the per capita incomes of the most developed countries. Adequate energy supplies are needed to improve living standards in less-developed nations.

Policy-makers now focus on energy, environment, and sustainability. For example, incentives are often used to reduce the environmental impact of energy use by increasing efficiency or substituting more environmentally benign energy resources for damaging ones. Such actions can make development more sustainable and improve living standards through a cleaner environment. Policies often reflect the concept that energy consumers share some responsibility for pollution and its impact and cost. Price increases to account for environmental costs have been implemented for energy resources in some locations.

Ecosystems are fragile and resources are scarce in many regions, and ecosystem protection requires that energy activities be carefully managed. Air, land, and water are being degraded in most areas, and life-forms such as mammals, birds, reptiles, plants, and aquatic life are threatened. Many of these concerns are associated with energy use, but in many countries energy options are limited. Energy is either imported using foreign exchange, which might also be used for purchasing items such as educational materials, medicine, or other development needs, or it comes from using local biomass. Large-scale consumption of biomass resources, however, leads to air and water pollution, deforestation, soil erosion, and global climate change.

The local culture can be viewed as part of the local environment. To be sustainable and responsible, development must be sensitive to its impact not just on the natural environment, but also on the local culture. Many argue that cultural diversity is as important to the planet's survival as biodiversity, but it is probably more endangered and less protected. Preserving cultures can be challenging because even though part of a culture is easily observed, its essence is often hidden and not shared with outsiders. Also, determining the "positive" elements of a culture and its development is subjective and based on one's values and experiences.

Energy and Society

Energy-use patterns in countries differ markedly. Globally, for instance, 20% of the population accounts for 70% of energy use. The United States is responsible for approximately 25% of total world energy consumption and 23% of greenhouse gas emissions. An average U.S. person consumes 230,000 Cal daily of food energy—115 times the 2000 Cal needed to survive. A typical western European uses as much energy as 80 people in sub-Saharan Africa. If the rest of the world were to use energy at the same rate as the United States, world energy-use would increase by about four times and the environmental impact would increase many times more than that.

There are many ways to reduce energy use. If 1% of the 140 million cars in the United States were tuned, gasoline consumption would decrease and 1 billion pounds of CO_2 emissions could be avoided. The United States could reduce its annual energy consumption by 50% by 2030 with efficiency increases and the use of renewable energy.[16] The use of renewable energy resources decentralizes the energy supply and allows for greater participation in energy decisions. For example, solar-related energy technologies offer greater local participation in energy decisions and have the flexibility to adapt to local conditions, costs, and benefits.

Increasing the efficiencies of energy systems can often improve living standards and personal fulfillment. Significant reductions in energy-use can in many instances be attained with little sacrifice of quality of life. An examination of 35 industrialized nations showed no correlation between energy-use and a wide set of social indicators (including life expectancy, literacy, unemployment, crime, suicide rates, and environmental quality indexes).[16] Another study showed that Sweden has a per capita GNP near to that of the United States but outranked the United States on almost every other social indicator while consuming 40% less energy per capita.[16] Some factors contributing to this difference include transportation variations (e.g., higher gasoline taxes, smaller cars, better public transportation, and geographic compactness in Sweden) and less wasteful commercial and residential energy-use. With changes in lifestyle, energy-use can be further decreased.

The effects of energy technology on society extend beyond energy issues because technology can generally influence social, cultural, and living standards. Industry is often regarded as a main contributor to societal well-being and the possession of technology is a source of societal prestige and identity. Technology often helps to integrate societies through shared resource-consumption patterns, values, awareness, and communications, but it can also stratify socially between the wealthy and the poor members of a society. The social consequences of deploying technology often depend on the social and institutional context. In some cases, technologies yield benefits that can be supported by social policies.

Energy and Sustainable Development

The 1987 Brundtland Report of the World Commission on Environment and Development defined sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs." This definition implies that actions of present societies should not threaten existing cultures and their evolution or future living standards.

The degree to which sustainable development can be achieved by countries varies because countries differ according to characteristics such as size, wealth, living standards, culture, and political and administrative systems. Wealth and advanced technology may make it easier for industrialized countries to strive for sustainable development. However, the reversal in the trend towards declining carbon emissions after the oil price decline in 1986 illustrates that this concept does not always apply. The basic motivations and desires of societies, countries, cultures, and people appear not to have changed, and these aspirations often require increasing energy use and often yield correspondingly increased emissions.

Transforming behavioral and decision-making patterns requires the recognition that current development paths are not sustainable. History suggests that such recognition occurs only when short-term consequences are obvious, as in the case of an "oil price shock" or a disaster such as a drought. To successfully mobilize the resources needed to reduce the risks associated with energy use and related environmental considerations, society must perceive the potential long-term consequences associated with present behavior patterns. Translating the future threats associated with continual increases in energy use and carbon emissions into immediate priorities is and will likely remain one of the most difficult challenges facing policy-makers.

POSSIBLE ENERGY-USE MODIFICATIONS TO IMPROVE LIVING STANDARDS

Solutions to energy problems that can improve living standards can be technical and nontechnical (e.g., reducing energy usage by changing lifestyles and increasing public awareness and education).

Energy Conservation and Increased Efficiency

Many energy-efficiency improvement and conservation measures have been applied, including regulations and standards for cars and buildings; incentives to stimulate investments in energy-conservation equipment; energy auditing and reporting procedures, especially for energy-intensive industries; and the promotion of relevant research and development.

Significant potential exists through improvements in energy efficiency for decreasing global energy consumption and enhancing the reliability of energy supplies and improving their longevity.[27–30] Improvements in energy efficiency often require modifications in energy-use patterns. Energy-efficiency measures can often be implemented quickly for devices with rapid turnover (e.g., light bulbs, refrigerators, cars). Power stations and similar infrastructures typically have much longer lifetimes. Despite high initial capital costs, efficiency measures can result over time in considerable economic savings for both individuals and societies. An example of the latter case is the elimination of the need for new power stations through high-efficiency electricity utilization.

Efficient energy use is particularly important to developing countries as it can forestall the need for large capital investments. Developing countries often lack financial resources and investment in efficient new technology is typically much less expensive than retrofitting old power plants. Expanding the economies of developing countries using modern technology allows them to bypass the inefficient technologies used in industrialized countries in the past.

Energy conservation also involves formulating appropriate energy pricing policies, using good maintenance and operation practices, and adopting efficient and effective load-management strategies. Significant reductions in consumer energy costs occur in many instances when appropriate energy-conservation measures and programs are implemented, and the payback period is often less than two years. Yet, energy-conservation programs are not undertaken on a significant scale in many countries.[30] Some reasons for this include:

- Technical difficulties (e.g., lack of reliable and efficient technologies)
- Managerial and institutional barriers (e.g., lack of appropriate technical input and program-design and monitoring expertise; inadequate program-management and training)
- Economic shortcomings (e.g., lack of financing mechanisms; inappropriate pricing of energy commodities)
- Inadequate information transfer (e.g., lack of information on technologies and related matters)

Achieving the potential gains associated with improved energy efficiency requires efforts by consumers, manufacturers, energy suppliers, and governments. Mechanisms are needed to encourage cooperation and overcome the potential obstacles to efficiency improvement. For example, incentives such as tax breaks can be provided to improve the efficiency of providing products and services. Incentives for the accelerated replacement and decommissioning of inefficient equipment can also be beneficial. Of course, practical limitations exist on increased efficiency due to factors like economics, sustainability, environmental impact, safety, and societal and political acceptability, and the desired balance among these factors often affects living standards and depends on a society's culture.

Strategic Planning

Addressing energy concerns while accounting for existing and desired living standards and culture requires long-term strategic planning. Otherwise, actions are likely to be inefficient, ineffective, and uncoordinated, and their potential benefits are not fully achieved. One strategic plan to address energy concerns is "New Earth 21," proposed in the 1990s as a long-term and comprehensive initiative to develop strategies that all countries can undertake cooperatively to address environmental degradation and achieve sustainable development.[31] The plan includes worldwide promotion of energy efficiency and conservation (within 10 years), the large-scale introduction of clean energy resources, including renewable and nuclear energy (within 20 years), and development of innovative energy technologies (within 50 years). Promotion includes increasing public awareness of the benefits of energy conservation through education and training and encouraging the development of comprehensive energy-conservation policies, particularly in areas of public welfare, transportation, and industry. The degree to which strategic plans prove to be acceptable or are able to be implemented in different countries depends in large part on cultures and living standards. However, the degree to which a country adopts such a plan also can affect its future living standards and cultural development.

ILLUSTRATION OF CONNECTION BETWEEN ENERGY AND SOCIETY/CULTURE

The country of China has presented in recent years one of the most notable examples of the importance of the relation between energy and societal living standards and culture. Energy use in China has been growing markedly between 2000 and 2005, especially in urban areas, leading to an increased standard of living. Improvements in living standards and greater affluence have led to changes in culture, including a trend towards greater consumption of resources and a stronger desire for consumer goods. This change in cultural behaviour can in turn fuel further increases in demands for resources, including energy. A spiral effect can thus develop, where increased energy use improves living standards and changes culture, which in turn leads to further increases in energy use.

Although there are many benefits to the population of China associated with these changes, there are also difficulties. Increased pollution has been one undesired consequence. Also, scarcities of material and energy resources have developed within the country and beyond. The demand for additional resources in China between 2000 and 2005 are thought by many to have significantly affected prices for commodities throughout the world.

CONCLUSION

Energy impacts a society's living standards and culture. This article explained this impact by considering energy use and its relation to population, the environmental impact of energy use, the impact of energy use on society, living standards, culture and sustainability, and possible energy-use modifications to improve living standards and sustainability.

ACKNOWLEDGMENTS

The author is grateful for the support provided for this work by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Colonbo, U. Development and the global environment. In *The Energy-Environment Connection*; Hollander, J.M., Ed.; Island Press: Washington, DC, 1992; 3–14.
2. Barbour, I. Energy. In *Technology and Society: A Bridge to the 21st Century*; Hjorth, L.S., Eichler, B.A., Khan, A.S., Morello, J.A., Eds.; Prentice Hall: Upper Saddle River, NJ, 2000.
3. *Technology and Society: A Bridge to the 21st Century*; Hjorth, L.S., Eichler, B.A., Khan, A.S., Morello, J.A., Eds.; Prentice Hall: Upper Saddle River, NJ, 2000.
4. Haberl, H. The global socioeconomic energetic metabolism as a sustainability problem. Energy **2006**, *31*, 87–99.
5. Niele, F. *Energy: Engine of Evolution*; Elsevier: Oxford, U.K., 2005.
6. Midilli, A.; Ay, M.; Dincer, I.; Rosen, M.A. On hydrogen and hydrogen energy strategies-II: Future projections affecting global stability and unrest. Renewable Sustainable Energy Rev. **2005**, *9* (3), 309–323.
7. Ausubel, J.H. Energy and environment: The light path. Energy Syst. Policy **1991**, *15*, 181–188.
8. *The Wiley Encyclopedia of Energy and the Environment*; Bisio, A., Boots, S., Eds.; Wiley-Interscience: Toronto, Canada, 1996.
9. Perman, R.; Ma, Y.; McGilvray, J. *Natural Resource and Environmental Economics*; Longman: London, 1996.
10. Rosen, M.A.; Dincer, I. On exergy and environmental impact. Int. J. Energy Res. **1997**, *21* (7), 643–654.
11. Sathaye, J.; Ketoff, A. CO_2 emissions from major developing countries: Better understanding the role of energy in the long term. Energy Int. J. **1991**, *12*, 161–196.
12. Scheraga, J.D. Energy and environment: Something new under the sun? Energy Policy **1994**, *22*, 798–803.
13. Speight, J.G.; Lee, S. *Environmental Technology Handbook*, 2nd Ed.; Taylor & Francis: New York, 2000.
14. Winteringham, F.P.W. *Energy Use and the Environment*; Lewis Publications: Boca Raton, FL, 1992.
15. Rosen, M.A. Energy considerations in design for environment: Appropriate energy selection and energy efficiency. Int. J. Green Energy **2004**, *1* (1), 21–45.
16. Rosen, M.A. Energy efficiency and sustainable development. Int. J. Global Energy Issues **2002**, *17* (1–2), 23–34.
17. Dincer, I.; Rosen, M.A. The roles of science and technology in energy and environment research and development. Int. J. Energy Res. **2001**, *25* (13), 1165–1187.
18. Dincer, I.; Rosen, M.A. Current and future perspectives on energy use and environmental impact. Int. J. Environ. Pollut. **1998**, *10* (2), 240–253.
19. Goldemberg, J.; Johansson, T.B.; Reddy, A.K.N.; Williams, R.H. *Energy for a Sustainable World*; Wiley: New York, 1988.
20. Strong, M.F. Energy, environment and development. Energy Policy **1992**, *20* (6), 490–494.
21. The World Bank *World Development Indicators 2000*; The World Bank Group: Virginia, 2000.
22. Brownsword, R.A.; Fleming, P.D.; Powell, J.C.; Pearsall, N. Sustainable cities—modeling urban energy supply and demand. Appl. Energy **2005**, *82*, 167–180.
23. *World Energy Outlook*; Organization for Economic Co-operation and Development and International Energy Agency: Paris, 1995.
24. Eden, R.J. World energy to 2050: Outline scenarios for energy and electricity. Energy Policy **1993**, *21*, 231–237.

25. Organization for Economic Cooperation and Development. *Energy: The Next Fifty Years*; OECD: Washington, DC, 1999.
26. World Energy Council. *Global Energy Perspectives to 2050 and Beyond*; World Energy Council: London, 1995.
27. Graedel, T.E.; Allenby, B.R. *Industrial Ecology*; Prentice Hall: Englewood Cliffs, NJ, 1995.
28. MacRae, K.M. *Realizing the Benefits of Community Integrated Energy Systems*; Canadian Energy Research Institute: Calgary, Alberta, 1992.
29. Wilbur, L.C. *Handbook of Energy Systems Engineering: Production and Utilization*; Wiley: Toronto, Canada, 1985.
30. Painuly, J.P.; Reddy, B.S. Electricity conservation programs: Barriers to their implications. Energy Sources **1996**, *18*, 257–267.
31. Okamatsu, S. MITI's centennial vision of global environment and technology and the response to global warming: Concerning new earth 21. In *Energy Politics and Schumpeter Dynamics*; Krupp, H., Ed.; Springer: Tokyo, Japan, 1992; 335–348.

Maglev (Magnetic Levitation)

Kevin C. Coates
Transportation & Energy Policy, Coates Consult, Bethesda, Maryland, U.S.A.

Abstract
As the fairly recent deployment of two different maglev transport systems in Asia has unequivocally demonstrated, the use of magnetic levitation for suspension and transportation propulsion is not only cost effective, but also introduces higher levels of system reliability not possible with other transportation technology. This entry provides a snapshot review of some of the more notable developments in this rapidly evolving transportation field.

INTRODUCTION

Magnetic levitation is a method of using the forces produced from either electromagnets or permanent magnets to suspend, support, guide, separate, or propel objects.

Transportation systems employing some form of magnetic levitation are known as maglevs and consist of vehicles moving along dedicated guideways. Using magnetic levitation as a means of locomotion represents a revolution in transportation because several inherently undesirable characteristics of wheeled transportation are eliminated or dramatically reduced—namely, friction (wear and tear), vibration, and noise (see Fig. 1).

Maglev technology is not "train" technology, and it is not compatible with any conventional railroad track design. Indeed, the scientific and engineering challenges of developing ultrasafe and high-reliability maglev ground transportation systems with top speeds comparable to turboprop and jet aircraft (500–580 kph) rival any of the world's great engineering achievements, including the world's most advanced space programs. For example, complex algorithms are used to control and operate maglev vehicles, and cost-effective construction techniques must be developed to build highly precise and extremely stable support structures known as guideways.

It should be emphasized that maglevs are complete transportation systems. The term maglev refers not only to the vehicles, but also to the vehicle/guideway interaction. Maglev system guideways and vehicles are precise design elements specifically tailored to each other for the creation and control of magnetic levitation.

There are several magnetic levitation approaches, all of which have their unique characteristics, advantages, and disadvantages. Due to space limitations, only the world's first three commercially available maglev systems will be discussed in this entry, along with the very mature Japanese high-speed system, still undergoing improvements at the maglev research and development (R&D) facility in Yamanashi Prefecture.

MAGLEV'S SOCIETAL IMPACT

Maglevs are not just exotic transportation technologies designed for high speeds; they are actually vehicles for societal change. For instance, the deployment of an extensive high-speed maglev network for electric-powered intercity transportation would significantly lower America's dependence on an increasingly unstable world oil supply. Use of lower-speed maglevs for commuter applications or for inner-city transit would also further lower oil dependence by coaxing people out of their cars for those longer point-to-point trips. These lower-speed systems also have the advantage of being nearly silent and vibration free, while able to operate safely on the steepest of grades even during inclement weather. Most important, these systems are designed to be safer than any other transportation mode ever invented because derailments are virtually impossible due to the way the vehicles fit around or within their guideways. In addition, braking requires no friction and therefore is unaffected by surface conditions (ice, snow, or rain).

In a November 2005 speech, U.S. Secretary of Transportation Norman Y. Mineta stated that America's cities were too far apart to justify a national rail system, such as in Europe or Japan. Maglev makes this an outdated statement.

Although this conventional wisdom may apply to conventional railroads, America's cities are not too spread out for a national high-speed maglev system that would be competitive with air travel. A high-speed maglev's top cruising speed is in excess of 500 kph (310 mph), and combined with very quick acceleration and deceleration

Keywords: Magnetic levitation; Transportation; Maglev; Transportation policy; Linear induction motor, (LIM); Linear synchronous motor, (LSM); Noncontact linear generator; Magnets; EMS; EDS; Transrapid; HSST; Linimo; MLX01.

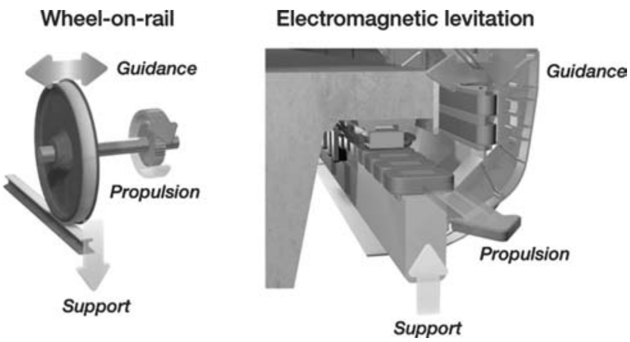


Fig. 1 Maglev eliminates friction for guidance, support and propulsion, graphic courtesy of Transrapid International, Berlin, De.

this makes it a perfect technology for travel distances between 50 and 1000 km (30–600 mi), especially when trip times, reliable operations, overall environmental impact, energy consumption, and safety are combined for consideration (see Fig. 2).

In addition, a 500-mile maglev line could make a few stops along its route to service those smaller communities that the airlines simply fly over. A one- or two-minute stop is all that would be required at each station. Several trains a day would reconnect these smaller communities to larger cities and dramatically discourage long, boring, slow, and energy-wasteful highway trips by car.

What Mr. Mineta failed to mention was that many American cities are too close together to justify financially inefficient short-haul air travel, yet the federal government continues to subsidize (i.e., encourage) these operations, which only add to the congestion of major airports while simultaneously imposing a disproportionately high cost in airport delays (see Fig. 3).

The most compelling questions are why America doesn't have a national transportation and energy policy, and why important transportation deployment considerations aren't being seriously considered by the U.S. government to provide relief to our serious congestion and transportation-related energy problems (see Fig. 4).

The following ten facts support the case for building a new high-speed, electric-powered, intercity maglev transportation system in the United States:

1. Half of all domestic air travel in the United States is for trips of 600 mi or less.
2. No large American city is farther than 500 mi from the next large city.
3. High-speed maglev trip times are competitive with door-to-door airline travel up to 600 mi.
4. Maglevs operate reliably in weather that typically grounds or disrupts air service.
5. America's transportation system is 96.9% oil reliant.
6. Jet aircraft are noisy and heavy air polluters, whereas maglevs are emission free (while maglevs do not spew emissions along their rights of way (ROWs), it is conceivable that the electricity generated to power them can come from less-than-green sources. That being said, a national energy policy dedicated to reducing or

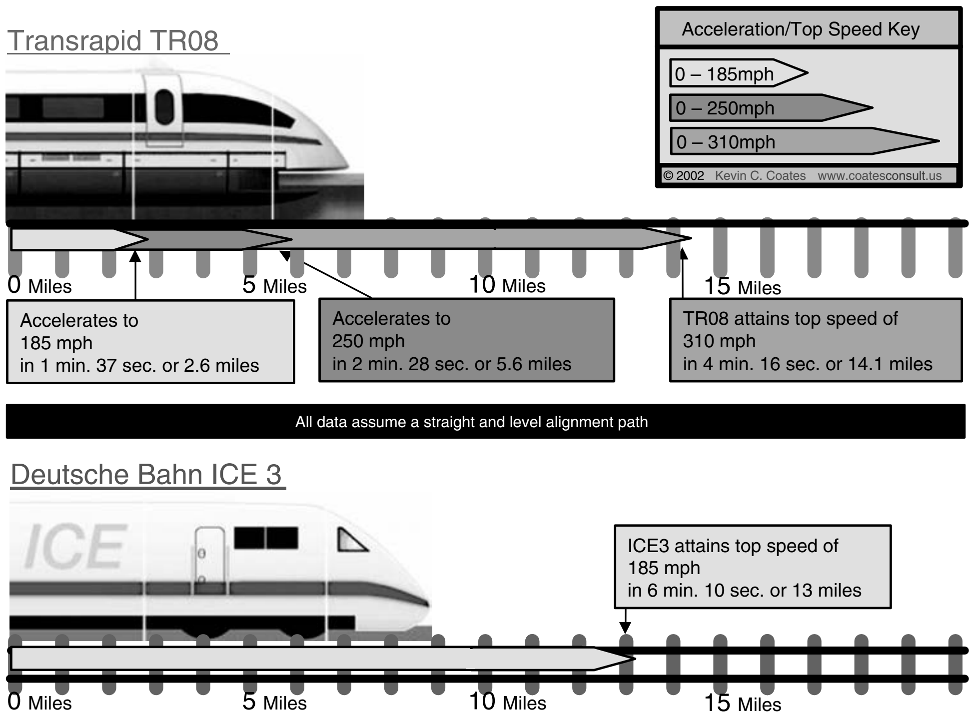


Fig. 2 Comparison between maglev and typical high-speed rail (HSR) acceleration & top cruising speeds.

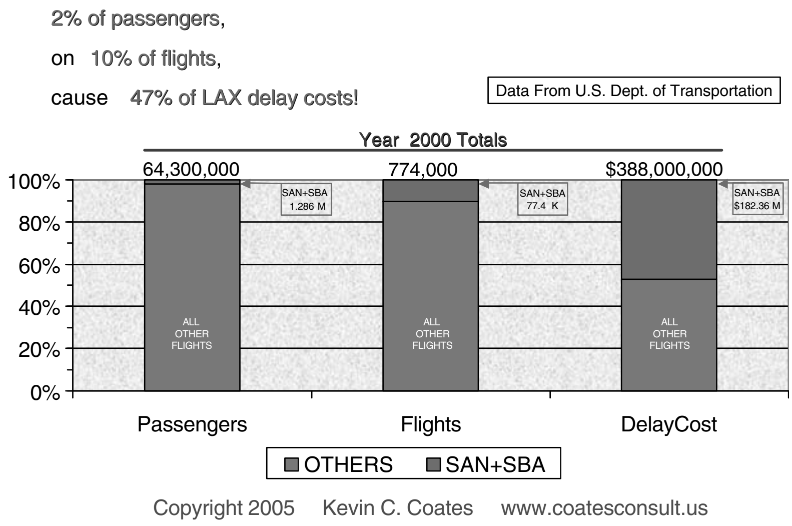


Fig. 3 Year 2000 LAX delays-the argument for high-speed ground transport between airports & outlying cities.

eliminating electricity generation plant emissions means that all electric-powered transportation becomes that much cleaner. Naturally, maglevs and other electric-powered modes are undeniably cleaner to operate along their ROWs than any combustion-powered mode).

7. Increasingly constricted world oil supplies are causing higher fuel prices that negatively impact U.S. transportation and the overall economy.
8. Higher fuel prices will inevitably give way to fuel shortages as world demand exceeds world production, further hampering domestic commerce.
9. America's 300 million people now consume nearly 22 million bpd of oil out of a 2005 world daily average of 86.6 million bpd, while China, a nation of 1.3 billion, consumes the second-largest amount of oil at 8.6 million bpd, meaning that China's explosive economic growth is driving rapid energy demand while high energy prices have yet to force U.S. citizens to realign their energy consumption patterns in any significant way.
10. Electric-powered intercity and commuter maglev networks could significantly reduce U.S. daily oil consumption and provide an effective alternative to hydrocarbon combustion-based transportation.

Given these hard realities, America will be forced eventually to reassess and reprioritize its national energy and transportation priorities. The longer America takes to address its growing national transportation and energy emergency, the more difficult the recovery will be. Words, arguments, discussions, meetings, studies, and the oil industry's propaganda and machinations by its political representatives will not prevent or reverse the physical realities of our "energy consumption crisis."

The drilling of thousands of oil wells now adds up to one big hole of dependency for America (and the world) to climb out of. As Will Rogers once quipped, "When you find yourself in a hole, the first thing to do is stop diggin'." Well, maybe America should heed ol' Will's advice and stop drillin', because our oil-based economy is simply not sustainable.

This brings up the real issue preventing maglev (or high-speed rail) deployments in the United States—the politics of vested interests.

TRANSPORTATION POLITICS

Maglev systems are now well-proven technologies. The difficulties preventing maglev deployment in the United States are certainly not technical. The main obstacle is the powerful political influence exerted by an existing oil-dependent transportation industry.

The need for a new national rail-type carrier is increasingly obvious to people familiar with America's looming energy crisis, but Amtrak is not the answer. Amtrak is a decrepit, dysfunctional, inefficient, unreliable, and expensive-to-maintain passenger rail system with an unworkable business model. To be fair, Amtrak was doomed to fail from the start. It was Congress that gave Amtrak an impossible national service mandate, while its former railroad owners assigned Amtrak passenger trains secondary status to freight rail traffic when riding their privately owned rails. The mere fact that Amtrak survived over 30 years is a testament to the public's demand and need for passenger rail travel. If this were not so, yearly Congressional Amtrak operational "subsidies" (as rail opponents prefer to call investments) would not have been dispensed.

The dissolution of America's passenger rail system did not happen by accident or in a political vacuum.

30 Potential Deployment Factors When Choosing Between Transportation System Modes	High-Speed Maglevs	Low-Speed Maglevs	High-Speed Rail	Commuter Rail	Subways	Light Rail/Trolleys	Bus Rapid Transit	Roads/Highways	Airlines &/or Airports
Capacity Flexible	H	H	H	H	H	M	H	L	L
High Capacity	H	H	H	H	H	M	H	L	M
Passenger Load/Discharge Rate	H	H	H	H	H	M	H	H	L
Hi-speed for Long Haul Travel	H	H	H	M	M	L	L	L	H
Severe Weather Impact	L	L	M	M	M	H	M	H	H
Safety If Top-speed Malfunction	H	H	L	M	M	M	M	L	L
ROW Footprint/Land Bifurcation	L	L	H	H	H	H	H	H	H
Operations Manpower Intensity	L*	L*	M	M	M	H	H	H	H
Initial Infrastructure Costs	H	M	H	H	H	H	L**	H	H
Life Cycle Costs	L	L	H	H	H	H	M	H	H
System Topographical Adaptability	H	H	L	L	L	L	L	L	L
Per Passenger Mile Cost	L	L	H	M	M	H	M	H	L***
Maintenance Costs	L	L	H	M	M	M	M	H	H
Operational Costs	L	L	H	M	M	M	M	H	H
Comfort	H	H	H	M	M	M	M	M	M
Air, Water & Soil Pollution Impact	L	L	L	M	L	L	M	H	H
Internal Noise	L	L	L	L	H	M	M	M	H
External Noise At Top Speed	M	L	H	H	H	M	H	H	H
Vibration	L	L	H	H	H	M	H	M	H
Transports Own Fuel Supply	N	N	N	Y	N	N	Y	Y	Y
Oil Dependent Operation	N	N	N	Y	N	N	P	Y	Y
Carries Own Propulsion System	N	N	Y	Y	Y	Y	Y	Y	Y
Overall Energy Efficiency	H	H	M	M	M	M	M	L	H
Operational Reliability	H	H	M	M	M	M	M	M	H
Disruption to Community	L	L	M	M	M	M	L	H	H
Fully Automatic Operation	Y	Y	P	N	P	N	N	N	N
At-grade Crossings w/Other Modes	N	N	Y	Y	Y	Y	Y	Y	N
Parts Availability	H	H	H	H	H	H	H	H	H
Prime Energy Flexibility	Y	Y	Y	N	Y	Y	Y	N	N
Urban Station Compatibility	H	H	H	H	H	H	H	L	L
Dual Mode	N	N	N	N	N	N	Y	N	N

KEY

COLOR CODED QUALIFIERS	QUANTIFIERS & THEIR SYMBOLS	
Very positive attributes	YES	Y
Average attributes	PARTIAL	P
Negative or low attributes	NO	N
Neutral or no deployment impact	HIGH	H
* Fully Automatic Operation	MODERATE	M
** Deployment Costs Soar With Dedicated ROW Construction	LOW	L
*** Huge Government Subsidies Distort Real Pass/Mile Cost		

Fig. 4 Coates' Maglev matrix. The overwhelming case for Maglev deployment.

During the first half of the 20th century, the decision to "invest" federally in the construction of a national network of airports—and to create a nationally funded and operated air traffic control system to manage the flight connections between them—put the first nail in the coffin for long-distance train travel in America. The Federal-Aid Highway Act of 1956 provided for a 65,000-km national system of interstate and defense highways. This act soon put the kibosh on most medium- and short-distance passenger rail travel.

Simultaneously, no money was being provided to the privately held railroads to shore up their aging infrastructure. As more money was being poured into the infrastructure for the other two modes, the railroad companies saw their passenger rail business drop off. Finally, they petitioned Congress to allow them to operate only as freight carriers and pawned off their passenger rail service on the federal government—hence, the creation of America's "national rail system," Amtrak.

However, after 50 years of heavy federal subsidies, America's highways and runways are increasingly more congested while oil supplies are increasingly less certain. America's first step to reduce oil dependence needs to be the shifting of the burden of intercity travel away from airlines/airports and cars/highways to more capacity-flexible and prime energy-flexible electric-powered rail or fixed guideway systems. The most technologically advanced of these systems is maglev.

With maglev systems in place, bustling intercity stations can be placed in downtowns. For instance, a downtown New Yorker could travel to the downtowns of Washington, DC; Boston; Baltimore; or Albany in about an hour and be on time—to the second—over 99% of the time. Travel between downtown Chicago and Manhattan would take approximately three hours, regardless of most weather conditions. This means, of course, that few people would ever fly these routes again, thus freeing valuable runway slots for longer-haul flights and obviating the need for expensive airport expansion projects. Indeed, the country's several ongoing, federally funded, multibillion-dollar airport expansion projects represent a logical potential source for future maglev funding ($20 billion was the total system cost estimated in the early 1990s for building a high-speed maglev system from Boston to Washington, DC), especially taking into consideration that at some point, rapidly rising fuel costs will eliminate airlines as mass-transit passenger carriers. Transportation planners need to be looking ahead at these very real limitations on the growth of hydrocarbon combustion transportation modes. Simply by transferring most of the present yearly federal subsidies for airport expansions to intercity maglev projects, America could begin to chart a path toward a more sustainable and reliable transportation system and significantly reduce its reliance on oil for transport. This is obviously not a technical hurdle but a political issue that needs to be understood from the perspective of being a national defense priority. If foreign oil supplies were suddenly suspended tomorrow, would Americans have a high degree of intercity mobility? No.

Yet soaring energy prices are not the worst problem facing energy consumers, for the high prices are merely a symptom of an imbalance in the supply-demand ratio. As world oil demand begins chronically to outstrip world production capabilities, not only will prices soar, but also, it is inevitable that shortages will eventually occur. The real problem is not just oil supplies, but America's overreliance on one prime energy source for 97% of its transportation. With electric-powered transportation, any number of "prime movers" can be used to generate electricity (see Fig. 5), making the country less vulnerable to selfish or corrupt political forces.

For most downtown travelers, to get to an outlying airport, check in, go through security, take the long walk to a gate, wait for boarding, board, taxi from the tarmac to a runway, take off, fly, land, taxi to a gate, unload, walk through an airport, grab a taxi or rent a car, and travel into another city takes a lot of time and energy. Without ever considering air travel time, downtown-to-downtown air travelers burn up over four hours just moving around on the ground. Downtown maglev stations would cut that ground travel time from hours to minutes and cut out much of the aggravation and stress in the process.

TECHNICAL APPROACHES AND CONSIDERATIONS

While some maglevs have top speeds in excess of 500 kph (310 mph), actual travel speeds would vary according to the route, just like short-haul flights. Efficient and comfortable high-speed air or ground travel requires extremely straight and relatively flat ROW's. The idea that America's Interstate Highway System ROW could host high-speed maglevs is not based on well-thought-out reasoning. While some extremely flat and straight sections of the interstate could certainly be used, most of the system would be able to share only small portions of its ROW because the highway undulations and curves were designed for comfortable top speeds of only 80 mph. If high-speed, energy-efficient operation is the goal, a high-speed maglev traveling four times faster than a car traveling on the highway at 70 mph would need a completely new and dedicated ROW over most of its route, especially between the end of the departing acceleration phase and the beginning of the final deceleration phase for arrival into a station.

Fig. 5 Electricity makes wind-powered high-speed maglevs possible, photo courtesy of Transrapid International, Berlin, De.

With a winding, undulating route, a maglev would be constantly accelerating and decelerating. As with all electric motors, the amps consumed during any acceleration phase are several times the amps required for operation at full speed. Additionally, because theoretically, all electric motors can become generators through the regenerative braking process, some high-speed maglevs could actually be designed to generate electrical power to the grid during a steady deceleration phase, thus improving our overall transportation system's energy efficiency.

There are several technological approaches for creating magnetic levitation that can be similar or very different, depending upon the manufacturer. For example, the world's first commercially deployed high-speed maglev in Shanghai, China, uses ferrous electromagnets to attract vehicles to the underside of a guideway (see Fig. 6), while onboard computers modulate magnetic strength to maintain a precise 1-cm gap from the sides and undersides of the guideway, resulting in the vehicles riding 15 cm

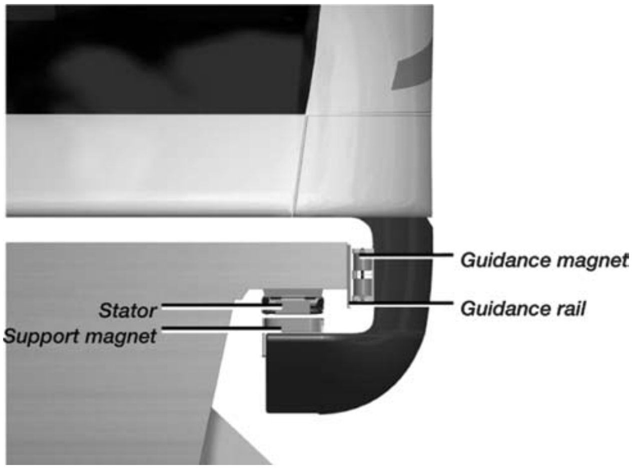


Fig. 6 Cross-section of maglev/guideway interface showing 1 cm magnet/guideway gap, vehicles are suspended 15 cm above guideway, graphic courtesy of Transrapid International, Berlin, De.

(about 6 in.) above the guideway during travel. This German-developed Transrapid high-speed maglev is called an electromagnetic suspension (EMS) system.

The second commercial maglev deployed this century is a low-speed HSST version (Linimo) in Nagoya, Japan, which is also an EMS, but its two key elements are the reverse of the Transrapid EMS system (see Fig. 7).

Electromagnetic suspension is essentially an electric motor broken down to its two key elements: the rotor and the stator. In a typical electric motor, a rotor is connected to the shaft that does the work. Rotation occurs when the windings in the stator surrounding the rotor are electrically "excited." Transrapid EMS maglevs are basically big electric motors laid flat across the landscape.

The low-speed EMS vehicles carry the stator function while the guideways serve the rotor function. This EMS approach is known as a linear induction motor

Fig. 7 Super-quiet low-speed (100 kph) "Linimo" maglev in commercial operation outside Nagoya, Japan.

(LIM)—hence, the Linimo name for the Nagoya system. The LIM approach allows for less expensive guideway construction but has limited practical top speeds—typically, around 200 kph (120 mph)—because increased speed necessitates a larger stator (vehicle) and adds undesirable weight and bulk that detract from high-speed performance. The LIM system also requires a contacting power pickup rail along its entire length, so it is not entirely a contact- or friction-free system, although propulsion and braking functions are achieved purely through electromagnetism and are free from friction.

As LIM system vehicles increase in weight and speed, guideways need to be sturdier, thus driving up guideway construction costs. In the mid-1970s, during the R&D of the German Transrapid prototype (the origin of the Japanese HSST design), it was decided that the need for speed made the LIM impractical because ever-larger vehicles would increasingly limit performance. Because of the Germans' desire to develop a high-speed intercity connector, they chose the "motor in the guideway" principle for their ensuing development. This was first demonstrated in public on the Transrapid TR-05 in 1979.

In the Transrapid TR-06, TR-07, Shanghai's TR-08, and the TR-09 now under development for a Munich airport connector, the vehicle acts as the rotor, and the guideway provides the stator function and easily supports any additional stator weight for producing higher speeds. This type of system is known as a linear synchronous motor (LSM).

With an LSM, passengers ride inside a vehicle (rotor) powered by an electrified cantilevered guideway (the stator). This arrangement allows for much higher speeds because vehicle weight tends to remain constant while static guideways carry the weight-intensive power delivery system for high-speed propulsion, suspension, and guidance. Along the entire length of the guideway, three-phase cables run through stator packs that are attached underneath the guideway cantilevers on both sides. While this arrangement increases initial construction costs over the LIM (motor in the vehicle) approach, it represents only 3%–4% of the guideway cost and makes very high speeds possible without any contact between vehicle and guideway. Indeed, the bulk of the guideway cost in Shanghai was for building an extremely stable structure to handle the speeds, loads, and potential seismic activity up to 7.5 on the Richter scale. At speeds above 80 km/h, power is delivered to the maglev vehicles through noncontact linear generators. Power is delivered via contacting power rails for lower speeds in and near stations. A series of onboard batteries provides redundant backup power to maintain vehicle levitation en route in case of any propulsion power failures.

To achieve and guarantee optimal ride comfort with minimum maintenance, the geotechnical challenges peculiar to maglev are formidable. These challenges include highly demanding deformation limitations,

Liq–Mobil

long-term stability of foundations under dynamic loads (including earthquakes), analysis of the entire foundation-support beam system, and optimization of the foundation systems for cost-effective design. Deformation considerations include immediate settlement, primary settlement due to consolidation, plastic settlements resulting from secondary consolidation or creep, total plastic settlements due to dead load, total settlements due to cyclical loads from vehicle operations, elastic settlements due to dynamic loads, and total anticipated settlement during operation.

For both LIM and LSM systems, vehicle levitation is achieved via onboard computer control units, which sample and adjust the magnetic forces of the onboard electromagnets as they are attracted to the underside of the guideway cantilevers (see Fig. 1 and 6). Vehicles move along the guideway with their cast-aluminum support arms wrapped around the top cantilevers of the guideway's I- or T-shaped cross section. The support arm's upward-facing suspension magnets are attracted to stator packs attached underneath the cantilevers on the LSM or the rails on the LIM guideway—designs that make derailment virtually impossible and maglev travel safe at any speed.

Regardless of load and speed, both LIM and LSM onboard control systems maintain a 10 mm gap with a $\pm$ 2 mm tolerance between the vehicle's support and guidance magnets and the guideway. Both systems are fully computer controlled and run automatically. These are not mechanical systems, such as wheeled trains or monorails, but digitally controlled and operated electronic transportation systems.

Central Japan Railway's MLX01 high-speed system, still in development in Yamanashi Prefecture, uses superconductor magnets to create magnetic repulsion to suspend its vehicles about 11 cm (about 4 in.) above the guideway (see Fig. 8). This is known as an electrodynamic suspension (EDS) maglev.

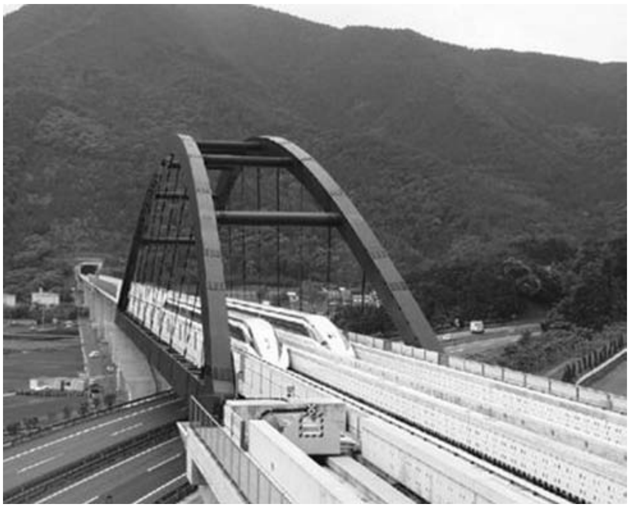


Fig. 8 Two Central Japan Railway MLX01 superconductor high-speed maglevs on test track in Yamanashi Prefecture.

RECENT MAGLEV HISTORY

To the lay person, it might seem logical, easier, and preferable that a magnetic levitation system would use repulsion magnetism. However, it was a computer-controlled magnetic attraction design, Germany's Transrapid TR-08, that was destined to be the world's first high-speed maglev deployed in a full-scale commercial project—in China (see Fig. 9).

The contract for the high-speed Shanghai airport connector was signed in January 2001. In less than three years, the 30-km (19-miles) airport connector commenced commercial operations. The speed of the construction project is as remarkable as the maglev's speed, especially when one considers that there was no maglev industry infrastructure in place in January 2001.

By the end of 2005, the Shanghai maglev airport connector had safely transported over 5 million passengers with an on-time-to-the-second—reliability level of 99.92%. In April 2006, after two years of normal commercial operation, the government's "expert group"—the Ministry of Science and Technology—as well as all the other affected agencies declared the maglev demonstration line an unmitigated technical and engineering success. The Ministry of Science and Technology then announced that the demonstration line would be extended 7 km farther, to the site of the 2010 World Expo and under the Huangpu River to the downtown Shanghai Railway station for an intermodal connection. From Shanghai, the line will then extend 163 km to the southwest to the resort city of Hangzhou, thus providing high-speed maglev access from Hangzhou to Pudong International Airport and a trip time of only 30 min for a distance of approximately 200 km (124 mi). Pudong

Fig. 9 A Chinese soldier salutes an arriving maglev as it enters Longyang Road station in Shanghai, photo by Kevin C. Coates.

Airport to downtown Shanghai will take only a comfortable 10 min by maglev and cost less than half the present fare for the hour-long taxi ride—and considering how harrowing Shanghai taxi rides can be, riding the maglev will be much safer.

While the Shanghai project was under way, another attraction maglev system was being built in Nagoya, Japan. This was the low-speed HSST, or Linimo, maglev. The Linimo began commercial operations in March 2005 to coincide with the start of the 2005 World Expo. This HSST 100 has a top speed of 100 kph (60 mph) and links 9 stations along a 9.6-km (5.6-mile) route. In the first three months of operations, the system transported over 10 million passengers with near-perfect (99.97%) on-time reliability and perfect safety.

The next maglev system scheduled to go online is in Daejeon, Korea, a city about 320 km (130 mi) south of Seoul. Built by the Korean company Rotem, this maglev system uses the same magnetic attraction methods as the Japanese HSST. The system is expected to be in commercial service by April 2007. The Chinese have developed their own HSST spinoff and are starting construction of new low-speed maglev lines.

Midway between Tokyo and Osaka, Central Japan Railway operates a 30-km maglev test track in Yamanashi Prefecture. The company's MLX01 maglev vehicle uses magnetic repulsion produced by powerful onboard superconducting magnets. As of late 2005, the MLX01 was still in development. However, it is the world speed-record holder for maglevs at 581 kph (360 mph). The plan is to extend the test track in both directions to provide additional high-speed service between Tokyo and Osaka and to provide relief to the corridor's heavily traveled Tokaido Shinkansen (137 million passenger trips in 2005), which now operates 12 trains per hour in each direction during peak travel periods.

Beside these advanced and mature maglev technologies, a result of some 40 years of development, there are a host of new companies around the world looking into different ways to achieve magnetic levitation, including the incorporation of permanent magnets into various design approaches. However, these systems remain in various stages of R&D.

THE MAGLEV PAYOFF

While "magnetic levitation" and "maglev" are becoming accepted parts of the national lexicon, the technology has yet to be deployed as a solution for America's transportation needs. This is in part because few policy-makers truly understand how the several variations of the technology work or, more important, how the domestic travel experience would be dramatically improved over its present ultrareliance on highway and air travel. And even if policy-makers are aware of the manifold benefits of maglev technology, their private political-campaign funding requirements seem to trump their national-interest concerns. How else can the lack of a single U.S. maglev system project under construction in 2005 be explained?

America has the need; it has the money; and it can buy and build the technology. What it clearly lacks is the political will to address its national transportation infrastructure problem in any meaningful way. The airline industry is failing miserably, requiring tens of billions of dollars in federal bailouts to keep planes in the sky. And state highway budgets throughout America are in the red because they cannot keep up with the cost of repairing an expanding road infrastructure without raising taxes. Yet only Amtrak is continually in the budgetary crosshairs.

In spite of the recent successes in Asia, which prove that maglevs are not "pie in the sky" dream machines but ultrareliable transportation engineering realities, American transportation policy continues to be wed to 1950s-era delusions of perpetually cheap gas and the seduction of open highways, although neither of which has been the case for a long time. America has simply not made the mental transition from seeing maglevs as promising transport modes of the future to the ultrareliable transporters they are now. Indeed, much of America has not awakened to the fact that its suburbs have transitioned into urban centers that need new transit options. Yet America's lack of any working maglevs is understandable, given that maglev technology was aggressively pursued, developed, and brought to maturity in far-off Europe and Asia. Meanwhile, American commuters are spending more time stuck in traffic.

CONCLUSION

Maglevs are merely the logical progression of the electricity revolution that was begun by Thomas Edison and Nikola Tesla in the late 19th century. The idea of using electricity to create a magnetic field is the basic premise behind both electric motors and generators. Indeed, maglev vehicles and their guideways are basically long electric motors when they accelerate, and they can function as generators when they decelerate.

Scarcely 100 years ago, when electricity began to be distributed into people's homes, it was viewed with fear and amazement, and it was not well understood (and still isn't by many people). However, it was not long before societies came to rely on the now-omnipresent supply of electric power to illuminate homes and factories in the evening, to cool and dry rooms in hot and humid climates, to power labor-saving machines safely, and to make instantaneous telecommunications possible worldwide. Indeed, it is our reliance on readily available supplies of reliable electricity that defines our world. Without it, our modern world would cease to be modern.

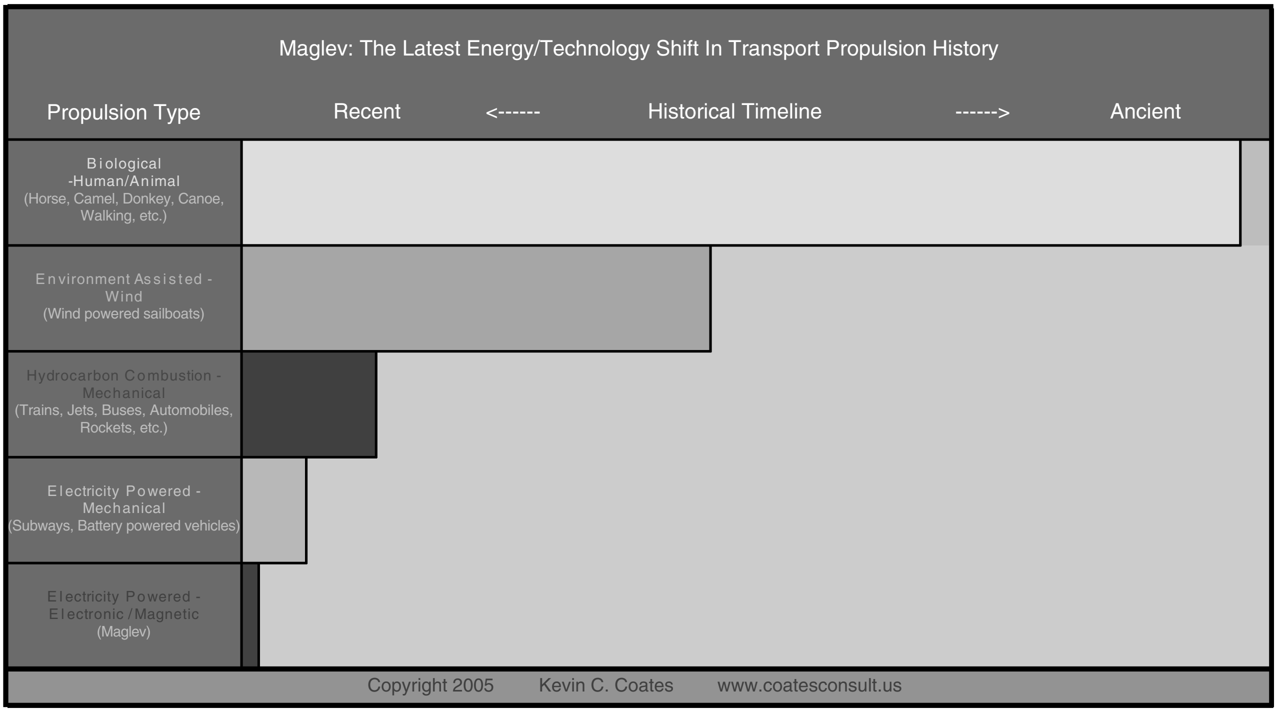


Fig. 10 Maglev: the latest energy/technology shift in transport propulsion history.

Although maglevs were conceived in the early 20th century, it was the rapid advancement in computer processing in the late 20th century that really propelled maglev development forward and transformed it into today's premier transportation option. Maglev transport is simply a logical next step in our society's electrical and transportation evolution (see Fig. 10).

As maglev systems continue to come online around the world, and as the price of oil continues to climb, questions surrounding this seemingly magical transport technology will naturally and inevitably arise. Questions surrounding energy efficiency; construction challenges; and the costs of system deployment, operations, and maintenance can all be answered and the engineering challenges overcome.

The answer to the equally important accounting questions is the impact maglev will and can have on society: the reduced desirability or need to fly between cities 600 mi or less apart, reduced air pollution levels, quieter and cleaner cities, lower national energy consumption, less use of land corridors to transport more people per hour than any other mode, and the ability to locate stations in urban environments. Clearly, the many tangential benefits of maglev transport transcend the seemingly black-and-white fiscal considerations.

In spite of many erroneous reports to the contrary, maglev systems are cost effective and fit seamlessly into the vision of developing sustainable and livable pedestrian communities that enhance, rather than compromise, citizen mobility.

Shifting from hydrocarbon combustion to electric-powered transportation promises to alter America's urban landscape dramatically for the better, improve the overall quality of life of its citizens, and improve America's national security. Indeed, it is maglev's promise of a more energy-efficient, cleaner, reliable, safer, quieter, and more sustainable transportation future that truly makes this technology so fascinating.

Management Systems for Energy

Michael L. Brown
Energy and Environmental Management Center, Georgia Institute of Technology, Savannah, Georgia, U.S.A.

Abstract
Organizational energy management has generally adapted a project-oriented focus, with success determined by the number of projects completed. While successful energy management projects must be part of a viable energy management program, organizations must also broaden their focus to enable energy management to adapt to constantly changing business environments and organizational objectives, achieve the desired results, and sustain those achievements. Following the model developed for quality and environmental management, a management system for energy based on the plan-do-check-act cycle of continuous improvement is presented and discussed in this section. A management system refers to the steps an organization takes to manage its activities in order for the products or services it produces to meet established objectives. The evolution of energy management practice, necessary elements of a comprehensive management system for energy, management system standardization and registration processes, and elements common to all management systems are also presented.

BACKGROUND

Traditional management practices sufficient to guide and direct static organizations are no longer satisfactory because today's business environment, production inputs, resource costs, and organizational objectives are constantly changing. The traditional management approach, where plans are made and then executed but not followed up, is inadequate in addressing these ever-changing situations. From an energy manager's point of view, these inadequacies mean that energy management projects may not be as successful as predicted, project results are not recognized, improvements are never incorporated into daily operations, and savings are gradually lost, or projects may never even be started due to inadequate resource support. A dynamic environment necessitates flexible management capable of achieving cultural change in the organization. To accomplish this, the simple "plan-do" approach to energy projects must be expanded to encompass a feedback loop that actually requires management involvement and support, and encourages and supports continual improvement.

To structure an organization for continual improvement and cultural change, the plan-do management plan, is superseded by the plan-do-check-act (PDCA) cycle. First proposed by Edward Deming as an approach to improving quality management practice, the PDCA cycle (shown in Fig. 1) is a four-step model for achieving organizational progress. Just as a circle has no end, the PDCA cycle is repeated over and over to achieve continuous improvement in the desired process.[1]

MANAGEMENT SYSTEMS

A management system refers to what the organization does to manage its activities so that the products or services it produces meet established objectives. Organizational objectives are set internally and can include the following:

- Satisfying the customer's quality requirements
- Complying to regulations
- Meeting environmental objectives
- Improving organizational energy efficiency

In a very small organization, there is probably no "system" as such, just a way of doing things. In most cases, this method is not written down, but exists in the head of the owner or manager. The larger the organization, and the more people involved, the greater the likelihood that there are some written procedures, instructions, forms, or records. These help ensure that everyone is not just doing their own thing, but going about organizational business in an orderly and structured way so that time, money, and other resources are utilized efficiently.

To be really efficient and effective, an organization must manage its methods by systemizing it. This ensures that nothing important is omitted and that everyone is clear about who is responsible for doing what, when, how, why, and where.

Management system standards provide the organization with a model to follow in setting up and operating its management system. This model incorporates the features

Keywords: Energy management systems; Energy master planning; ANSI/MSE 2000 standard; Management systems for energy; Organizational energy management.

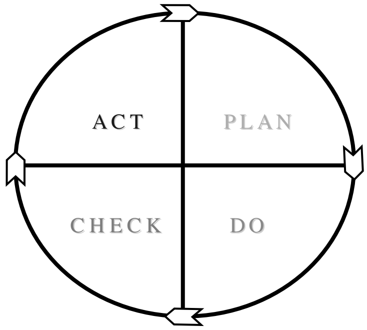


Fig. 1 Plan-do-check-act cycle for organizational change.

that experts in the field have collectively deemed state-of-the-art. A management system which follows the model or conforms to this standard is built on a firm foundation of state-of-the-art practices.[2]

The PDCA management cycle sets the stage for a system to manage organizational change and it serves as the process followed in all management system standards. The "plan" element of a management system involves recognizing a problem or opportunity and planning the actions necessary to solve the problem or take advantage of the opportunity. Planning often includes collecting, reviewing, and analyzing relevant data in order to identify a problem.

After the plan is complete, it must be implemented. This is the "do" step in the PDCA cycle, and it involves acting on the plan—preferably following a clearly defined process. The "check" step involves assessment and may include monitoring and measuring. Did the action achieve the planned change? When assessing the results of an energy management project, metrics defined in the planning phase are used to quantify the degree of change and detail the extent to which the project did or did not achieve the desired results. Checking is not limited to projects, but encompasses all phases of the management system, and checking activities are defined by a transparent process.

The final step in the process is "act." Here, the organization must take action based on what was learned in the "check" step. If the desired result was not achieved, the cycle repeats with a revised plan. If the planned result was achieved, most organizations tend to stop at this stage. With a management system in place, the organization goes further to incorporate improvements throughout the organization by developing or revising processes or procedures. This success forms the basis for planning new actions to yield even greater improvements.

Types of Management Systems

The PDCA management system can be applied to any organizational process in need of improvement. Management system standards have been developed for quality, environmental, energy, food safety, information security, global supply chain security, and social responsibility management. The management systems for each of these areas have been standardized and adopted as international standards by the International Organization for Standardization (ISO) headquartered in Switzerland. All of these international standards ensure standard processes and results that support international and national commerce.[3]

With worldwide adoption of other ISO standards, movements to evolve from national energy management standards to the process for development and adoption of an international standard is progressing rapidly.

THE STANDARDIZATION PROCESS

The American National Standards Institute (ANSI) has served in its capacity as administrator and coordinator of the United States private sector voluntary standardization system for more than 80 years. Throughout its history, ANSI has maintained the enhancement of global competitiveness of U.S. businesses and the American quality of life as its primary goal through promoting and facilitating voluntary consensus standards and conformity assessment systems and promoting their integrity.

ANSI facilitates the development of American National Standards (ANS) by accrediting the procedures of standards developing organizations (SDOs). These groups work cooperatively to develop voluntary national consensus standards. Accreditation by ANSI signifies that the procedures used by the standards body in connection with the development of ANS meet the Institute's essential requirements for openness, balance, consensus, and due process. At year-end 2003, there were about 200 ANSI accredited standards developers and there were more than 10,000 ANS.[4]

ANSI promotes the international utilization of U.S. standards, advocates U.S. policy and technical positions in international and regional standards organizations, and encourages the adoption of international standards as national standards where they meet the needs of the user community. The Institute is the sole U.S. representative and dues-paying member of the two major nontreaty international standards organizations, the International Organization for Standardization (ISO) and, via the U.S. National Committee (USNC), the International Electrotechnical Commission (IEC). As a founding member of the ISO, ANSI plays a strong leadership role in its governing body, while U.S. participation via the USNC is equally strong in the IEC.[4]

Although each country's procedures for developing and adopting national standards vary, U.S. procedure depends on the development of a management system framework by an ANSI accredited standards developer. The process includes review and voting by the full range

Table 1 Key elements of management systems

Element	Description
Clear-cut responsibilities	Explicit descriptions of management authority and expectations
Documented processes	Information on what is to be managed and how
On-going training	Promote awareness to ensure proper functioning of management system
Internal checks for conformance	Verification of conformity with standard through documented proof
Corrective/preventive action	Process followed to address problems or non-conformities within the system
Management reviews	Regular review of management system results by top management
Continual improvement	Along with sustained improvements, continual progress is a primary objective of management system implementation

of stakeholders. In the case of the ANS for energy management (originally developed by the Georgia Tech Energy and Environmental Management Center), stakeholders included well-known experts from associations related to energy and/or water; commercial energy users; consultants from energy engineering, energy technology, or management systems; educators (nonprofit) providing training in energy and/or water conservation and management; energy service companies serving as energy brokers or providers of energy services; equipment suppliers related to energy or water use; government (federal, state, or local) related to energy policy, research, and conservation; manufacturers with significant energy use in operations; regulatory agencies responsible for oversight of energy/water supply, storage, use, or disposal and utilities that provide energy or water service.

The demand for an energy management standard has been growing steadily around the world and there is a push for further national standards, a European Union standard, and an international standard. ANSI/MSE (management system for energy) 2000 was the first national energy management standard in the world, but Denmark, Ireland, and Sweden have also developed energy management standards.

ISO—a nongovernmental organization—is a federation of the national standards bodies of 149 countries from all regions of the world, including developed, developing, and transitional economies. Each ISO member is the principal standards organization in its country. Members propose a new standard, participate in its development, and provide support in collaboration with ISO Central Secretariat through the more than 3000 technical groups that actually develop or merge different versions of national standards. ISO governs the development process of a proposed standard in partnership with the broadest possible base of stakeholder groups, its adoption by this group using specific consensus procedures, and the process of ensuring equivalent application of the standard within business, government, and society. ISO standards are widely respected and internationally accepted by public and private sectors. The United States is represented in ISO by the ANSI.[5]

Key Management System Standard Elements

To truly accomplish permanent organizational change, the management system approach must be institutionalized through several common key standard elements. A list of key elements is presented in Table 1. Whether the management system is for quality, environmental, or energy management, it must incorporate these key elements to be successful and sustainable.

The first key standard element in any management system is clear-cut responsibilities. Using manuals and procedures, the responsibilities and authorities of each participant are explicitly defined. This allows every participant involved with the management system to know and understand their own and others' roles.

The second key element is documented processes. Processes controlled by the specific management system, as well as the methods utilized, are documented. Documentation can take numerous forms, but must ultimately be understandable to the users. The particular form of documentation is open and can include written text, figures, charts, photographs (digital or conventional), video tapes, or any combination of these media sources. The documents can be presented in a paper or an electronic format, but must be easily accessible to all users.

Ongoing training is essential to satisfy requirements for employee awareness and enable understanding that results in the proper execution of responsibilities. The training component of a management system must define the specific training needs for each position and list specific courses, registrations, or certifications needed to comply with the requirements. Generally, training requirements and individual completion records are retained for documentation purposes. While this sounds like a huge burden, most organizations combine these requirements with their ongoing training and documentation process, which are already in place.

Operating management systems are routinely subjected to internal checks for conformance (also referred to as internal audits). During the internal audit, the management system is examined by organizational employees to ensure conformity with the standard. Internal audits serve at least two purposes. First, audits familiarize organizational employees with the management system standard and the related processes and procedures used by the organization to ensure conformity. For day-to-day oversight, the process uses trained internal personnel to identify nonconformities within the management system standard rather than relying on external auditors. Equally important, the check will identify the successes and best practices related to the system. By using internal resources, the organization can correct nonconformities without the risk of losing third party certification to the standard.

Corrective/preventive action is another key component of an effective management system. Corrective action is initiated to eliminate the cause of a problem traced back to a nonconformity. A corrective action will, by definition, eliminate the root cause of the nonconformity. Because it addresses the cause, preventive actions also get to the root of a problem. Unlike corrective actions (which deal with existing problems), preventive actions are concerned with eliminating a potential problem (one that could occur, but has not been witnessed yet). Both corrective and preventive actions contribute to the reality and culture of continuous improvement in an organization.

Management reviews are essential because they constitute the ACT element of the system and are a key to achieving continual improvement. Management reviews are conducted at defined intervals, and involve a review of system progress and success, as well as documented actions to deal with failures. Top managers must participate in the review to see the progress achieved, demonstrate their commitment to the management system, and ensure resources are available for necessary projects and the implementation of results. The management review allows top managers to look at management system performance during the review period and take action on system failures.

Management systems share the common objective of continual improvement in the managed organizational function. ISO management systems include express requirements that the system and specific processes also demonstrate continual improvement. The demonstration must show that the organization has developed and sustained a culture of continual improvement through the management system. Defined metrics are used to document the progression of improvement achieved by the management process.

Documents Defining a Management System

Management systems are defined by families of standards that are usually referred to under a single, generic title for convenience. The primary standard defining a management system is referred to as a specification standard. The specification standard establishes the management system requirements that must be met and is used by auditors to verify that all requirements have been satisfied. A specification standard provides the organization with a model to follow in planning, implementing, and operating a given management system. This model incorporates the features on which experts in the field have reached a consensus as to the standard representing or incorporating state-of-the-art best practices. A management system which follows the model, or conforms to the standard, is built on a firm foundation of these best practices.[6]

In addition to the specification standard, families of management system standards also include guidance standards, often referred to as guidance documents, which provide additional explanations or address specific issues related to the specification standard. The ISO 14000 family of environmental management system standards contains guidance documents presenting general guidelines on principles, systems, and supporting techniques (ISO 14004), the environmental assessment of sites and organizations (ISO 14015), general principles on environmental labels and declarations (ISO 14020), principles and framework of life-cycle assessments (ISO 14040), and others. The entire ISO 14000 family contains 26 standard documents, including the specification standard, guidance standards, technical reports, and draft standards.[7]

Registration, Certification, and Accreditation[8]

It is the specification standards in the ISO management system families of standards that are used in third-party certification and registration. Certification refers to the issuing of written assurance (the certificate) by an independent, external body that has audited an organization's management system and verified that it conforms to the requirements contained in the specification standard. Completing registration means that the auditing body then records the certification in its client register. When these actions occur, the organization's management system has been both certified and registered. For practical purposes, the difference between certification and registration is not significant, and both are acceptable for general use.

Accreditation refers to the formal recognition by a specialized body—an accreditation body. Accreditation means that a certification body, generally referred to a registrar, is competent to carry out certification audits in specified business sectors. Certificates issued by accredited certification bodies, known as accredited certificates, may be perceived in the market as having increased credibility. An organization can decide to fully implement a management system yet elect not to conduct a registration audit with an accredited auditor. If this is accomplished, the organization can still have a functioning management system, but will not have certification from an independent registrar, which, in some cases, can offer a competitive advantage.

The registration process for a management system is not a one-time action. Once a management system is formally registered, the registrar will normally conduct annual or semiannual surveillance audits. Furthermore, registration certificates are issued for a fixed period of time, usually three years. To maintain the registration, a recertification audit must be conducted at the end of the registration period.

EVOLUTION OF A MANAGEMENT SYSTEM FOR ENERGY

Of course, the standards of interest to energy managers are related to energy management systems. The evolution to an adopted management system for energy standards has taken a long time to develop, and has involved numerous interim phases, as illustrated in Fig. 2. Prior to 1973, there was no formal system to manage energy. Because the cost of energy in the United States was so low, there was no incentive to even consider formalizing energy management. Only the most intensive energy users had any type of program, and this was generally minimal.

The initial Arab oil embargo by OPEC in 1974 caused a sharp increase in petroleum prices, making industrial energy conservation worthwhile. By the late 1970s, many companies were bringing in outside consultants, such as Georgia Tech, to direct energy conservation efforts. This was a logical solution, as the problem was new and no internal expertise has yet been developed.

As energy costs remained high, more companies recognized the value of energy management and began to train and develop internal energy managers. This was preferred over outside consultants because a company employee as energy manager would always be on-site and thus have far greater understanding of the manufacturing process and organizational decision making. The Association of Energy Engineers aided this effort by developing a certificate process for energy managers. When an employee satisfied the requirements of a Certified Energy Manager (CEM), a company could be assured that the individual was trained in many of the technical aspects of energy projects.

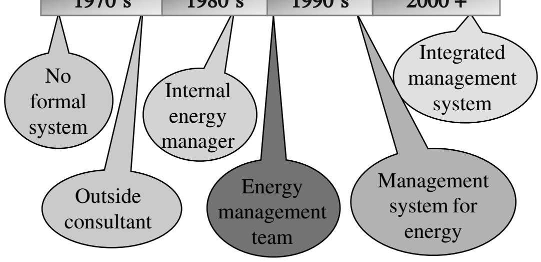


Fig. 2 Evolution of energy management.

By the early 1990s, most leading-edge companies had recognized the need for an energy management team. Because the management of energy crosses organizational boundaries, an effective energy management program cannot be restricted to one department. The team approach incorporates individuals from all areas within the organization involved with energy purchasing, company operations, production issues, facilities, engineering, maintenance, environmental management, and possibly others. By breaking down the "silo" mentality, energy management broadens the program's experience base and improves communication processes.

In 1998, Georgia Tech Energy and Environmental Management Center (EEMC) personnel utilized more than 30 years of experience with energy and extensive experience with ISO 9001, the quality management system standard, and ISO 14001, the environmental management system standard, to formulate a management system for energy. The primary driver for this effort was the need to sustain improvements from energy management projects instead of following the typical degradation of savings. The MSE 2000 standard was adopted as the U.S. national standard in April 2000. The ANSI/MSE 2000 standard document describes the essential elements of an effective energy management system. From a simple (but complete) system in a smaller organization to the sophisticated system implemented by large energy users, the standard is flexible enough to allow each organization to determine how they will implement necessary elements. Because the standard is modeled after other management systems, the structure for an auditable protocol was already available in order to allow registrars approved for ANSI/MSE 2000 to conduct third-party audits and certify organizations with formal registration.

Because there can be a number of different management systems in use by a given organization, the final step in the evolution of management systems will be an integrated system. An integrated system will permit shared management systems elements, such as internal auditing, corrective/preventive actions, and management review, in a single, unified system. Dissimilar elements remain separate and distinct, maintaining the unique strengths of each management system. While management system integration is a pursued objective, it will probably not be fully developed or widely implemented until sometime in the next decade.

CURRENT ENERGY MANAGEMENT SYSTEM STANDARDS

As discussed previously, management systems in general and management systems for energy in particular must follow a defined approval process to be adopted as international standards. The normal development process for a management system standard is for one or more national standards to be developed first. As international interest in energy management grows, an international technical committee is convened, and development of a consensus standard ensues. In most cases, the original national standard or standards serve as the starting point for development of an international standard.

The First European Conference on Energy Management was held November 23 and 24, 2005, in Milan, Italy. Because of changes in world energy markets as well as evolving legislation affecting energy use patterns, energy decision makers decided there was a global need for common and shared references. Due to the emphasis on the environmental impact of energy, European standards so far tend to emphasize this aspect, while ANSI/MSE 2000 integrates both technical and management aspects with an emphasis on energy, as well as environmental costs. As a consequence, Comité Européen de Normalisation (CEN) and Comité Européen de Normalisation Electrotechinique (CENELEC) decided (with the participation of many European National Standard organisations and global stakeholders, including the Georgia Tech Energy and Environmental Management Center) to establish a joint advisory group, CEN/CENELEC-BT/JWG-EM, on the subject of energy management. The aims of this joint working group are:

- Investigating and listing the existing and proposed laws and regulations in each European country and the European Union;
- Identifying and listing energy management standards, whether existing or under development;
- Defining the needs for one or more new standards based on current and future legislation and market requests; and
- Making recommendations on the standardization work to be undertaken.

The conference featured presentations on the four existing energy management standards (in order of their adoption as national standards):

- U.S. Std ANSI/MSE 2000:2005, Management System for Energy
- Danish Std DS 2403 E-2001-08-17, Energy Management—Specifications DS/INF 136 E-2001-12-0, Energy Management—Guidance on Energy Management
- Irish Std I.S. 393.2005 (12.8.2005), Energy Management Systems—Requirements with Guidance for Use
- Swedish Std SS 62 77 50 (2.10.03), Energy Management Systems—Specifications

ANSI/MSE 2000:2005: ORGANIZATION AND ELEMENTS[9]

The ANSI/MSE 2000 Management System for Energy is a standards document that describes the elements necessary to establish and maintain an effective strategic energy management system in a variety of organizations. The management system described by the standard is based on the Deming PDCA management cycle and can be used to establish a management system that actively promotes continual improvement.

All U.S. standards developers are required to review and potentially revise each standard adopted by ANSI at five-year intervals. The MSE 2000 standard was initially adopted in April, 2000, and a review by the standards developer (Georgia Tech Energy and Environmental Management Center) resulted in a revised document reflecting a process structure, which maintained compatibility with ISO 9001 (quality management) and ISO 14001 (environmental management). The revised energy management system standard was then reviewed by a canvas board of national energy management experts, resulting in a consensus vote and unanimous approval of the draft standard. ANSI formally adopted MSE 2000: 2005 in July 2005.

Process Format

During the five-year existence of ANSI/MSE 2000:2005, the originators of the standard, engineers from the EEMC recognized a number of difficulties in comprehension and implementation the standard. Many of these difficulties can be attributed to the standard's organization. Although any management system uses the PDCA process, the 2000 version of MSE 2000 was not organized into these categories. In fact, the standard was organized by activities, such as Energy Monitoring and Measuring. When organized in this manner, an element of the standard can contain any part of or all of the PDCA subprocesses, as shown in Fig. 3.

Because the elements of the standard contain numerous process activities, the standard authors decided that reorganizing the standard along process activities would

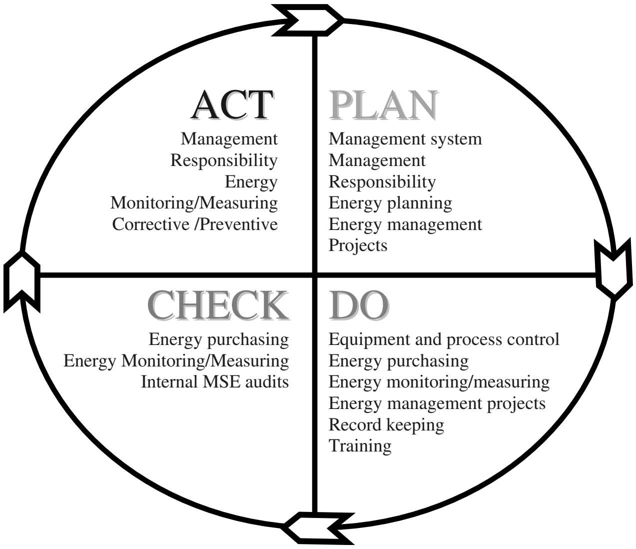


Fig. 3 MSE 2000 original standard organized into plan-do-check-act process.

be preferred. Fig. 4 illustrates the 2005 process-oriented standard.

Standard Elements

Table 2 describes the elements of the revised standard, ANSI/MSE 2000:2005. The standard is clearly organized according to the PDCA cycle, as shown by elements 6.0, 7.0, 8.0, and 9.0. In addition to the PDCA segments, the standard also contains two elements (4.0 and 5.0) that must exist to support these activities. Element 4.0, Management System for Energy, contains requirements necessary to support the general management system, including documentation, recordkeeping, and other general activities; while 5.0, Management Responsibility, details the attention and support required for successful energy management projects.

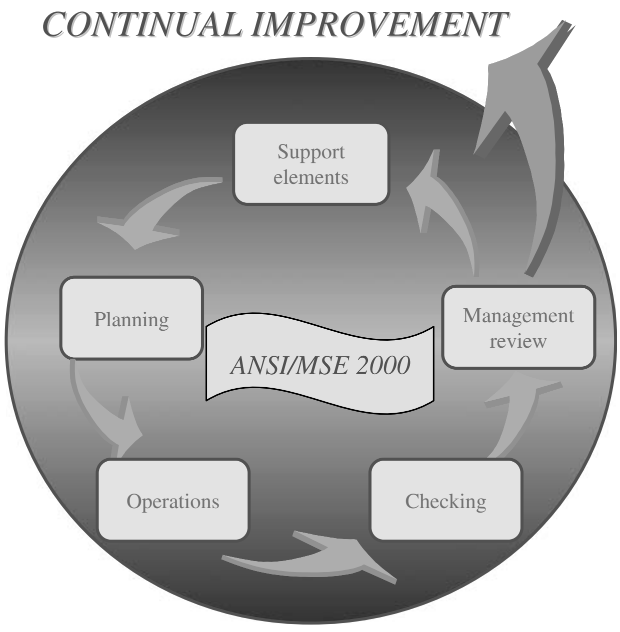


Fig. 4 ANSI/MSE 2000:2005 now reflects a process approach.

Due to increased concern, the development of new approaches or protocols, or to clarify or emphasize certain elements, subsequent additions were made to the content of the standard. These additions include: requirements for strategic planning; utility tracking; selection of significant energy uses and key performance indicators; more detail on energy purchasing, equipment, or building commissioning; energy management project selection and full project management; and control of outsourced energy services.

The involvement of executive or top management of the organization is critical if an effective management system is to be created and maintained for the long term. Management must demonstrate their commitment through providing manpower, physical resources, and funding. Management sets the tone of the system through the creation of an appropriate energy policy and strategic plan, choosing personnel responsible for operating the management system, and giving selected personnel the authority to make the system work.

Strategic planning and key performance indicators (KPI) were added to the standard to ensure complete alignment with the needs of executive management. Top managers look at an organization's key performance indictors to gauge performance, so including KPIs as a highly visible part of the energy management process makes the MSE 2000 system more meaningful to these decision makers. By including a strategic planning element in the energy management system, executives can identify and develop a process to meet the organization's energy, operational, and financial needs.

Because energy monitoring and purchasing must be considered in both the planning and doing phases of management, requirements were split between two elements. Energy monitoring and measuring was included in the standard under element 7.0, DO, and utility tracking was included in Section 6.0, plan.

During review of the standard, developers recognized that two areas, commissioning and outsourced services, had been overlooked. A growth of interest in these areas demanded their inclusion. Section 7.2.1, Commissioning, Recommissioning, and Continuous Commissioning, addresses commissioning in new, old, and maintenance situations. Section 7.4, Control of Outsourced Energy Services, requires that outside organizations participating in energy management be held to the same system requirements as internal staff.

Table 2 Elements of the ANSI/MSE 2000:2005 standard

Standard element	Content
4.0, Management system for energy	
4.1	General requirements
4.2	Documentation requirements
4.3	Recordkeeping requirements
5.0, Management responsibility	
5.1	Management commitment
5.2	Energy policy
5.3	Strategic planning
5.4	Authority and responsibility
6.0, Energy management planning (PLAN)	
6.1	Energy profile
6.2	External information
6.3	Energy assessment
6.4	Goals and targets
7.0, Implementation and operation (DO)	
7.1	Purchasing
7.2	Facility, equipment, and process control
7.3	Energy management projects
7.4	Control of outsourced energy services
7.5	Commissioning
7.6	Training, competence, and awareness
8.0, Checking and evaluation (CHECK)	
8.1	Energy monitoring and measuring
8.2	Internal MSE audits
8.3	Corrective and preventive action
9.0, Management review (ACT)	

MANAGEMENT DRIVERS[10]

A fundamental question regarding ANSI/MSE 2000 or any other management system is: What factors induce an organization to apply such a standard? The implementation of a structured, well-defined management system for energy has been shown to offer an organization numerous benefits.

Higher Energy Management Performance

Traditionally, energy has been treated with a crisis approach. Crisis energy management tends to focus on quick decisions and often high-risk solutions to the problem. ANSI/MSE 2000 uses planning and energy monitoring to define organizational objectives, to appreciate both the internal and external energy picture, and to improve the management of energy projects and results. Additionally, the standard results in better-trained employees in all areas critical to efficient energy usage. In all cases, improved energy management efforts will increase an organization's ability to stay competitive.

Greater Organizational, Involvement, and Competency Concerning Energy Issues

The MSE standard stresses a team-based approach and adequate training to build organizational competencies. The team concept increases involvement, widens the knowledge base of managers, and incorporates inputs from diverse perspectives. Requiring training of the employees involved with energy and energy management greatly improves their ability to manage this critical resource.

Increasing the number of persons involved in energy management within the organization improves the understanding of energy issues in every department, not just in those directly related to energy. The standard focuses on the importance of simple housekeeping measures (doing the small things consistently and correctly) to improve energy efficiency before considering capital projects. The team-based approach encourages all employees to develop ideas for improving energy efficiency, with generally surprising results. This creates a shared sense of purpose and esprit de corps that can increase involvement and improve morale.

Better Communication Regarding Energy Management both Inside and Outside the Organization

ANSI/MSE 2000:2005 requires that energy management performance parameters be communicated to team members, organization executives, and other employees who influence energy. Effective communication is mandatory to accurate planning and decision making. Furthermore, the implementation of an ANSI/MSE 2000 management system sends the message outside the organization that the organization is serious about energy management and dedicated to its efficient usage.

One of the cornerstone principles of management systems like ANSI/MSE 2000 is the collection, analysis, and retention of critical energy performance indicators. Consequently, energy managers are more articulate in conversations with other departments, resource suppliers, regulators, and neighbor facilities. It also means that energy managers and other personnel are aware of potential interactions with other systems, whether operational or organizational.

Improved Energy Efficiency, Reduced Costs, and Greater Consistency in Focus

ANSI/MSE 2000 emphasizes improving purchasing practices, operating performance, and equipment maintenance over intensive capital investment. Focusing on these low-cost, low-risk options improves organizational energy efficiency thus reducing energy costs, extends the life of energy assets, maximizes organizational value of projects, and optimizes the impact of considered capital investments. Documentation of the management system increases the consistency of purchasing, operating, and maintenance practices, and of organizational planning and decision making.

Although in some organizations energy costs may represent a small percentage of the cost of doing business, every organization expends significant funds for energy, and costs continue to rise. The potential for economic benefits should be identified in order to demonstrate to stakeholders the value of good energy management to the organization. The energy management system also provides the organization with the ability to link goals, targets, and energy management projects with specific financial results, thus ensuring that organizational resources provide the most benefit in terms of finances, resource management, and environmental impact.

In a relatively short period of time, the organization will develop and implement efficient management tools for defining corporate energy priorities and individual responsibilities. The system allows managers to better prioritize and defend their energy program resource needs. The ANSI/MSE 2000 system provides both team members and at-large workers a better understanding of what each is required to do and the means to do it consistently, competently, and efficiently.

Better Relationships with Suppliers

In a sector fraught with concern over rising energy prices and market restructuring, the energy monitoring, measuring, and profiling required under MSE 2000 provide detailed insight to an organization's energy needs, usage patterns, and incremental costs. This allows energy users to accurately define energy needs, which improves the ability of suppliers to develop accurate bid proposals and operating projections.

A management system for energy is a set of management tools and principles designed to guide the allocation of resources, assignment of responsibilities, and ongoing evaluation of practices, procedures, and processes that an organization needs to integrate energy issues into its daily business practices. By making energy management a part of daily functions, an organization avoids the trap of managing energy by crisis, relying instead on proactively planned responses. Effective energy management through implementation of ANSI/MSE 2000 yields the expected resource and cost savings, and additionally, it establishes a context of continual improvement that both sustains the gains made and puts the organization in a position to forge even greater energy efficiency and cost savings.

COMMON MANAGEMENT PROBLEMS ADDRESSED BY ANSI/MSE 2000

Implementing a management system defined by a documented standard, such as ANSI/MSE 2000, addresses at least ten common management-related energy problems faced by most organizations. In addition to addressing these recurring issues, an energy management system provides a flexible structure for sustaining the gains made in organizational energy management practice.

Lack of Organizational Communication

One common situation in most organizations is the lack of communication between different departments. Because energy is purchased in one department and consumed in another, and because energy systems are maintained and operated by still another department, the responsibility for energy is spread among so many that, effectively, no one is in charge.

The energy management team required by MSE 2000 is composed of representatives from all functional areas concerned with energy (e.g., purchasing, operations, engineering, and maintenance). Because all areas participate in the management of energy and adhere to a fully communicated structure, coordination is promoted. Instead of responsibility being lost between departments, the energy management team captures input from all areas to arrive at workable solutions to any problems or opportunities.

Insufficient Commitment from Executive Management

Because management has a bottom-line perspective, an energy manager situated in a single functional area can seldom obtain the attention of organizational executives. Energy is often viewed as a small potatoes issue and it is thus undeserving of much attention or resources. Another hindrance is that many of the issues competing with energy for attention (such as environment and safety and health) have legal mandates that force compliance and demand action.

Implementing a management system for energy overcomes this limitation by requiring upper management commitment to the development of an energy policy, the provision of adequate resources to the energy management program, regular review of energy goals and results, and the pursuit of continual improvement. This broad-brush approach changes the focus from a single department to the whole facility and brings credibility to energy management efforts. To achieve registration to the ANSI/MSE 2000:2005 standard, upper management must be committed to the energy program and demonstrate this commitment through participation and support.

Lack of Focus and Shifting Priorities

A lack of focus and shifting management priorities are common problems encountered by energy managers. Although the energy manager may have a clear focus, organizational management usually faces quickly changing priorities. Energy, safety, occupational health, environment, quality, productivity, process improvement, and cost controls are just a few of the conflicting priorities that can influence decision makers. With such a laundry list of objectives, it is no wonder that priorities become confused and shift so often. A recent solution to this dilemma is the advent of institutional management systems like the management system for energy. All of the specialized management systems grew out of the need to manage scarce resources, but still achieve the goals of the organization.

Reliance on a Single Person

While CEM certification through AEE indicates an individual has a strong technically based expertise in energy, it does not imply management specialization in either the individual or the organization. Even facilities with CEM-certified energy engineers rely on a single technical specialist for energy solutions. While this individual is without question well trained, this method puts the organization at risk because all of its energy management expertise is lost if that person leaves.

Another disadvantage is that the energy manager approach depends on the point of view of a single person, while MSE 2000 creates an energy management team and generates ideas from many different people with different perspectives. The team concept required by ANSI/MSE 2000 avoids the single point of view engendered by reliance on only the energy manager, who is generally part of the engineering department. This necessarily limited perspective usually focuses on how difficult an energy system is to install, maintain, and operate. Energy purchasing practices and system efficiencies are seldom considered, and selection of capital projects may result in unanticipated problems for other departments or systems. Including all functional areas in the decision process changes the perspective from a departmental to a facility-wide one, ensuring comprehensive planning.

Ineffective Follow-Through on Plans

Energy programs are initiated with the best intentions, but too often the momentum is lost and the organization returns to previous standard operating practices. An integral part of ANSI/MSE 2000 is a provision for continual improvement achieved by regular investigation of opportunities, the encouragement of employee input on potential projects, and acting on the information discovered. Elements of the ANSI/MSE 2000 standard that relate to continual improvement are energy planning, energy monitoring and measuring, and corrective/preventive action.

Energy planning involves using the current energy profile, external data sources, and assessments of equipment and processes to formulate savings opportunities. Energy monitoring and measuring refers to the collection of operational data to indicate energy problems or opportunities and the use of appropriate metrics to measure energy project success. Over time, the cumulative results of successful energy projects will be reflected in improvements in key performance indicators. Problems with the management system itself are mitigated through corrective/preventive action. In summary, ANSI/MSE 2000 establishes a permanent management structure that promotes follow-through and continual improvement in energy efficiency.

Lack of Institutional Procedures

One of the strong points in favor of ANSI/MSE 2000 is that it establishes institutional procedures to govern energy decisions. Without a system in place, decisions often seem haphazard. Energy purchasing and energy project decisions are not made the same way each time, creating spotty, inconsistent results in energy management.

The organizational policies associated with the energy management system are formalized in the energy manual. Energy procedures follow energy policies. The existence of well-defined and widely communicated energy policies and procedures provides a stable framework and promotes reliable, consistent results.

Continually Addressing the Same Problems

A common problem observed in any organization is the failure to conquer many small, recurring energy-related problems. Because there is no systematic approach to energy management, the same problems occur over and over, with energy engineers never seeming to get ahead enough to achieve real improvement. Much of this can be attributed to a lack of management support and the absence of management system structure.

Rising energy costs at a facility create a crisis management situation where other problems may be temporarily superseded by concentration on energy issues. Resources devoted to solving high energy costs, including addressing neglected maintenance and operating practices, may achieve real improvement. However, with no energy monitoring in place, it may be difficult to determine if real gains have been made, but any drop in the total cost for energy is attributed to the energy crisis management effort. Unfortunately, after some improvement is achieved, top management is quick to move on to more pressing problems and these improvements are then lost. Implementing an ANSI/MSE 2000 management system requires an organizational commitment to continuous improvement, implying the institution of a management structure wherein gains are maintained and new opportunities are pursued.

Limited Resources for Effective Energy Management

Inadequate resources can often be the death of an effective energy management program. Without exception, resources in the form of capital, labor, and raw materials are necessary to execute a program of resource management. To ensure adequate resource allocation within the ANSI/MSE 2000 system, management commitment is required not just in the form of system approval, but in the form of financial and manpower assets. Top management support not only assures that energy management will be an organizational objective, but also results in a firm commitment to the provision of resources adequate for an effective energy management system.

To maximize the impact of resources, the priority for energy management projects is the best return for the least investment. Fig. 5 presents the recommended opportunity hierarchy. Before concentrating on new capital equipment, ANSI/MSE 2000 requires a commitment to proper purchasing, operating, and maintenance practices. The base elements of the opportunity pyramid—purchasing, operation, and maintenance—create a stable foundation on which new capital equipment investment can rest.

Lack of Useful Energy Data

Complete and timely data is critical in optimizing the management process. Often, organizations make energy management decisions based only on bottom-line costs.

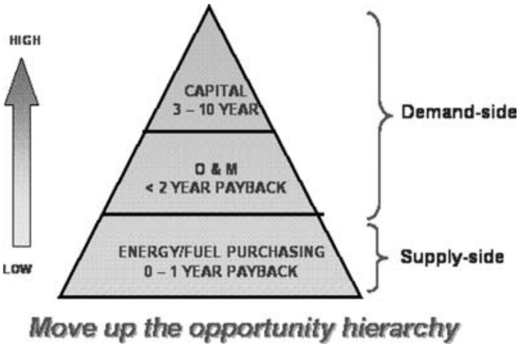


Fig. 5 Return on investment follows the opportunity hierarchy.

The ANSI/MSE 2000 standard contains strict requirements for energy data to be collected, analyzed, and retained. While external energy costs and internal consumption are important, normalized data indexes must be selected to yield accurate comparisons. The compilation of appropriate energy data is required for a complete (and thus effective) ANSI/MSE 2000 management system.

Ineffective Response to Changing Energy Markets

The deregulation of energy suppliers has greatly complicated the purchase of energy resources. The increased volatility of energy prices and uncertain availability of adequate resources in the future demand that organizations develop a workable plan to deal with frequent changes. The requirement for formal procedures specifying information, data, and processes for effective energy purchasing covers an area that is frequently neglected. Proactive approaches to purchasing decisions must be both defined and followed, with the management system ensuring conformance.

CASE STUDY: ANSI/MSE 2000 IMPLEMENTATION AT C&A FLOORCOVERINGS[11]

Energy management has long been considered a technical problem—that is, approached with capital projects to incorporate new technology. Plant engineers and/or maintenance personnel generally propose a project and upper management participates only in the approval or disapproval of a particular project. Usually, there is no direct connection between the business objectives of the company and any specific project.

C&A (formerly Collins & Aikmann) Floorcoverings is a major carpet and flooring manufacturer with five plants in the Dalton, Georgia area. This industrial sector is highly energy intensive, and the company is concerned with both increasing energy costs and environmental impacts. The company decided to further their company commitment to energy and environmental stewardship by implementing ANSI/MSE 2000, a management system for energy. Adoption of this standard publicly proclaimed company commitment to the business goals of reducing energy costs, energy consumption, and environmental impact (through the reduction of air emissions related to fossil fuel use). Because of the company's well-known emphasis on environmental responsibility, environmental improvement from the management system was the primary selling point for management.

The importance of these business objectives to the company served as a precommitment from upper management. That alone virtually guaranteed the success of the implementation. In addition, management decided to formally register the ANSI/MSE 2000 standard so that its use and results could be used as a marketing tool to emphasize C&A Floorcovering's position on energy and environmental issues.

Why ANSI/MSE 2000?

ANSI/MSE 2000 is an ANS encompassing a total approach to energy management. It covers both technical and management aspects of a system to manage and control the purchase, storage, use, and disposal of energy and water utilities. The standard provides structure for the system, but allows the flexibility to develop an appropriate and relevant approach.

As Fig. 6 illustrates, a major advantage of the system is that it aligns energy management projects with company goals and objectives. This ensures that the projects undertaken are carefully selected to help fulfill business objectives and provide maximum direct and indirect benefits. The result is a system that engenders full management support and the sustainability of project savings.

Based on a 2002 annual energy bill of almost $2 million and NO*x* emissions of over 10.6 tons, company management felt that ANSI/MSE 2000 could not only identify, but more importantly, sustain substantial savings. While management firmly believed that they had already implemented every significant energy project, they found that their savings tended to disappear over time, as operations inevitably reverted to previous behaviors and conditions.

To management's surprise, within months of implementing the system, employees were recommending energy management projects with significant impact. The energy and environmental manager credited increased communication, cooperation, and visibility for the creativity observed in these projects. Results from energy management projects (both savings and improvements in processes and procedures) continue to accrue, while procedures ensure new problems and opportunities are addressed.

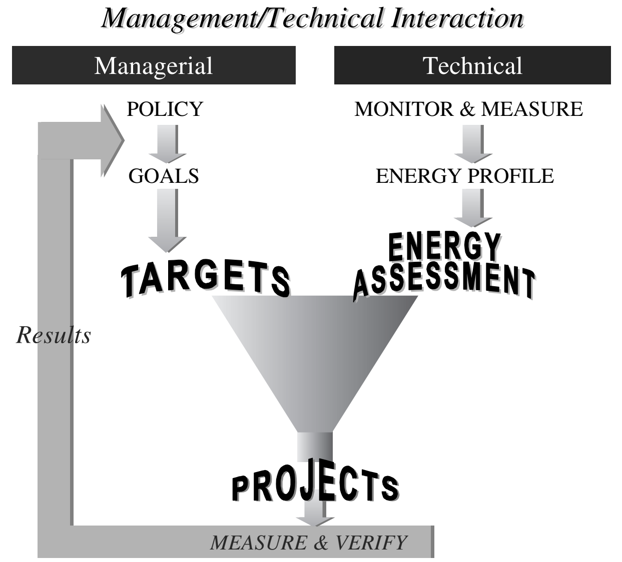


Fig. 6 ANSI/MSE 2000 provides full integration between management and technical aspects of the system.

INITIAL STEPS IN THE IMPLEMENTATION PROJECT

C&A Floorcoverings teamed with the Energy and Environmental Management Center (EEMC) within the Georgia Tech Enterprise Innovation Institute for the ANSI/MSE 2000 implementation project. Working with personnel from EEMC, C&A Floorcoverings personnel determined the first steps to take. Throughout the implementation project, EEMC provided courses and coaching designed to assist in a smooth and successful implementation.

Because resources were not available to implement ANSI/MSE 2000 in all five plants at once, the Energy and Environmental Manager decided to focus on the two largest plants. The Yarn and Dye plant is 162,000 ft^2 and the Finishing plant is 250,000 ft^2. Both include a variety of highly energy-intensive equipment. Because resources within the two plants were tight, C&A Floorcoverings and EEMC developed a scope and timeframe for the project that reflected these business realities.

Gained Commitment

The first requirement for beginning implementation was to gain broad support for the project, especially from upper management. This required educating upper management on the elements and benefits of the standard and tailoring this information to emphasize its support of corporate environmental and cost-savings goals. Support for the project was top-down, which encouraged support from managers and supervisors within the plants. C&A Floorcoverings also developed broad support by tailoring the message to address issues and concerns of various stakeholder groups.

Selecting the Energy Coordinator

As required by the standard, upper management appointed the energy coordinator, Kent Benson, and determined that he would also lead the implementation effort. It was important that the energy coordinator be in-house and familiar enough with plant operations to work well with the MSE Team. He also had to meet the other standard requirements for the energy coordinator, including having the necessary skills and training and possessing the responsibility and authority to function within the system.

Liq–Mobil

Table 3 C&A floorcoverings' energy and environmental objectives

Produce a high quality product using as few virgin raw materials as possible, including energy
Produce as few disposable by-products as possible, including direct and indirect emissions from energy usage
Lower production costs by using less energy

Selected the MSE Team

The first approach to assembling the MSE Team was to use an already existing group, but maintenance people on that team had no additional time for the implementation project. The reasons for the ANSI/MSE 2000 requirements of a broad-based team and necessary resource support (including time) became clear. By June 2001, a team of ten people from upper management, engineering, maintenance, purchasing, and technical services personnel were appointed as the MSE team. Because of the tight resources within the plant, members worked on the implementation as they could and most of the day-to-day responsibility fell to the energy coordinator. The major lesson learned was to ensure that resource allocations meet the planned timeline of the implementation. Here, C&A Floorcoverings chose a slow implementation due to resource constraints.

Determining Where We Stand on Energy

At the beginning, the requirement to develop an energy profile of the facility was perceived as a huge engineering project. It quickly became clear, though, that the energy balance and energy profile were much easier than expected. Software provided by Georgia Tech EEMC made developing a database of energy information simple; the program automatically prepares charts and graphs related to the data for trend analysis and informational uses. The energy balance used available information on energy consumption and the expertise of operators and maintenance personnel to fill in the holes.

Details of the energy balance were then used to develop a list of significant energy users in the plant. At first, the MSE Team tended to include far too much data in the significant users list. Later, it was pared down to a workable list consistent with Pareto's 80/20 rule (20% of the equipment will account for 80% of the energy used). The lesson learned here was to choose significant equipment wisely, as this list was later used to help select potential energy management projects. To ensure meaningful communication, the team selected appropriate energy indicators to assist in evaluating progress.

Developed Initial System Purpose and Metrics

As required by the standard, upper management participated in developing an energy policy. This policy serves as the reflection of management concerns and energy and environmental objectives (Table 3).

Once upper management had approved the energy policy, the MSE Team began developing initial energy goals. These goals, listed below, served to align energy management projects with management and business objectives. Implicit in the energy reduction goals are a reduction in associated emissions.

- To reduce the consumption of electricity per unit of product on an annual basis at each plant.
- To reduce the consumption of natural gas per unit of product on an annual basis at each plant.
- To reduce the consumption of water per unit of product on an annual basis at each plant.

Based on the elements of the energy policy, the significant energy users list, and knowledge of plant operations, specific targets (equipment or energy systems) were developed. Reductions in cost per unit of product and emissions per unit of product served as the metrics used to determine if each energy goal was reached. Once goals and targets were established, monitoring and measurement, the energy profile, and the energy assessment all provided data on the current status of the system and potential opportunities. A structured process was used to analyze potential opportunities and select those best matched with the targets and showing the best combination of benefits and costs. Part of the process was to evaluate the results of each project. As part of the system, these results are fed back into the system and result in an update of current goals or spur the development of new goals.

DEVELOPING THE MANAGEMENT SYSTEM

With the basic components of the management system for energy in place, the MSE team began developing system and program documentation.

Writing the Energy Manual

Using the guide provided, the energy coordinator developed policies appropriate to each element of the standard and wrote the C&A Floorcoverings' Energy Manual. While the energy coordinator wrote the first draft of the manual, all of the MSE team members participated in reviewing it and providing feedback on the ramifications of the policies, as well as any necessary changes. The Energy Manual served as the guiding document for the entire energy management system for all five plants.

The body of the manual is only nine pages, allowing flexibility for operational differences within the plants.

Developing Procedures and Work Instructions

To tailor the system to the particular facility, the implementation team formalized or developed written procedures for those operations affecting energy cost, consumption, or disposal (including emissions and waste streams). Initially, this was intimidating, as the energy team felt they would have to write procedures for the entire plant. Later, the number of procedures was pared down by concentrating on significant energy users and those processes necessary for operations. In a number of cases, the team had sample procedures or other documents (such as the MSE Procedures Guide) as guides. For clarity, each procedure was formatted to include information needed for document control and revision. Work instructions were prepared only when necessary, such as in an area with significant worker turnover. The total of all procedures and work instructions for the largest plant stand less than one inch high. In addition, many of the system procedures are applicable to all five plants and do not have to be redeveloped.

Implementing Procedures and Programs

As the methods to address each element of the standard were selected and the procedures and work instructions were developed, the team implemented each program in turn. Because the company was anxious to move into the core of the system (energy management projects), the energy team first developed the procedure for completing an energy assessment. EEMC had already provided the Energy Assessment Workbook, which guided an experienced staff member through the process of identifying potential opportunities for savings.

Next, the MSE team used the structured system presented by EEMC to develop their program for energy management projects. This approach ensured that energy projects would address the business objectives and goals of the company and that any changes made as a result of a project would be incorporated into system documentation and training.

The corrective and preventive action system provided a method to track and effectively solve any problems with either operations or the management system for energy itself. In an atmosphere of continual improvement, staff members report problems on a form, which is then used to identify the approach to be taken to solve the problem and check the effectiveness of the solution once it has been implemented. It also identifies whether any changes are needed in operational procedures or MSE documentation.

The training system pertains only to personnel who significantly affect energy purchasing, storage, usage, or disposal. Thus, the system is not unwieldy; it includes a training needs analysis and documentation that the training was completed. Training can range from on-the-job training (OJT) to training sessions as part of a larger meeting on formal education. Because so many of C&A Floorcoverings' employees are veterans and are already well trained, the MSE team found relatively few instances where training need assessments were extensive. But the consideration given to the needs assessment also fed directly into job descriptions and new-hire qualifications.

Once a year, C&A Floorcoverings holds a formal management review meeting. This meeting includes upper management and enables them to remain appraised of progress toward goals through energy management projects. It also affords an opportunity for them to examine results from internal audits of the system and review corrective and preventive actions.

The internal audit system concerns the management system for energy itself, rather than operational questions. Six internal auditors examine the system on a scheduled basis throughout the year to verify continued conformance with the standard, check implementation, determine the effectiveness of the system in meeting business objectives, and identify opportunities for improvement. The company's audit schedule calls for a one-half day audit of particular elements of the system four times a year. Findings from the audits are presented in the management review meeting and are addressed through the corrective action system.

SYSTEM TRANSFORMATION OF THE WAY PROJECTS ARE HANDLED

From stand-alone projects developed by single individuals, C&A Floorcoverings developed an approach that ensures that energy management projects are:

- Aligned with management policies and goals
- Appropriate to the plant and the problem
- Effective in solving the problem or making the improvement
- Reflected in permanent changes to operational policies and procedures
- Considered in revised goals and targets

Energy management projects address both opportunities and problems. Opportunities usually arise during the energy assessment or from an analysis of the appropriateness of a new technology. Problems surface through the corrective and preventive action program. In each case, the process ensures that only the most beneficial projects are selected for implementation.

Too many times in the past, a project would be completed only to find that the baseline or operational data to judge whether the project was successful was lacking. The MSE system ensures the tracking and evaluation of each energy management project. In order to complete this evaluation, it is crucial that appropriate metrics be selected early in the planning stage. Regular monitoring and measurement provide much of the data needed for project evaluation. In each case, the energy coordinator (as head of the MSE team) verifies the effectiveness of the project after implementation as well as the resulting operational changes.

The energy management system also ensures that any operational or policy changes are fed back into the system's documentation. The document control system is electronic, so the most current procedures and work instructions are always available. By referring to the training needs assessments, the MSE team knows exactly which operators are affected by the change and can immediately schedule them for appropriate training.

Management results are disseminated to all stakeholders, but are always meaningful to the individuals. This means that upper management concentrates on the overall effect on the energy indicators, technical personnel obtain more detail on specific projects that concern their areas, and operational personnel see the energy and environmental results and receive appropriately detailed information on any operational changes from the project. Program results against the baseline energy profile are communicated through posters visible throughout the plants.

Through the structure of the ANSI/MSE 2000 system, C&A Floorcoverings ensures that all energy management projects directly address the business objectives of the company. Business objectives, energy policy, energy goals, targets, and projects are consistently aligned. Regular management review meetings provide maximum visibility of results for management and the communication plan ensures that meaningful information is provided for personnel throughout the company.

SUMMARY AND CONCLUSION

C&A Floorcoverings has successfully implemented a flexible tool, ANSI/MSE 2000, for the management of energy. The system clearly provides consistent, relevant, sustainable results from the energy management projects selected.

For example, because ovens are significant energy users within the carpet industry, the initial plans concentrated on developing energy management projects for these target users. First, a project to reduce the burner chamber temperature on the regenerative thermal oxidizer at the recycling facility was selected. The oven had been operating at 1525°F, but analysis by operators determined that the temperature could be reduced to 1200°F with no increase in VOC emissions or visible opacity. Reprogramming costs were just under $10,000, but resulted in annual savings of $9500 and 1600 mcf of natural gas. While the simple payback is one year, the company is confident that the savings will be sustained over the coming years, as appropriate operating changes have been incorporated into operations.

Next, the MSE team examined the curing ovens in the precoat range at the finishing plant, which was operating at 325°F. They experimented with lower temperatures and settled on 270°F, a 55° decrease. This saved $14,100 and 2400 mcf per year, with minimal implementation costs. An added benefit of the lower temperature is longer conveyor belt life.

During the development of work procedures, the MSE team (which included operators) discovered that the thermal oxidizer on the vinyl curing range was running at 1400°F. Because opacity was the only emissions concern, the MSE team experimented with lower temperatures until 950°F was determined to be the ideal operating temperature. Annual savings are $55,000 and 9200 mcf of natural gas. An indirect benefit is that internal components will last much longer with this 450° temperature reduction and maintenance costs should be lower.

In the 24 months since implementation of the management system, C&A Floorcoverings has seen an additional 10% reduction in natural gas usage and cost. The company anticipates this trend will continue into the future.

REFERENCES

1. Tague, N.R. *The Quality Toolbox,* 2nd Ed.; ASQ Quality Press: Milwaukee, Wisconsin, 2004; 390–392.
2. Generic Management System Standards, http://www.iso.ch/iso/en/iso9000-14000/understand/basics/general/basics_3.html, International Organization for Standardization, 2006.
3. New Management System Standards, http://www.iso.org/iso/en/iso9000-14000/msstandards/index_five.html, International Organization for Standardization, 2006.
4. Introduction to ANSI, http://www.ansi.org/about_ansi/introduction, Washington, DC, 2004.
5. International Standards for a Sustainable World, International Organization for Standardization, 2005.
6. ISO 9000 and ISO 14000 in Plain Language, http://www.iso.ch/iso/en/iso9000-14000/understand/basics/general/basics_4.html, International Organization for Standardization, 2006.
7. The ISO 14000 Family of Standards, Guides and Technical Reports, www.iso.org, International Organization for Standardization, 2006.
8. Certification, Registration and Accreditation, http://www.iso.ch/iso/en/iso9000-14000/understand/basics/general/basics_5.html, International Organization for Standardization, 2006.
9. Brown, M.; Key, G. *Revisions to ANSI/MSE 2000, The Energy Management System Standard*, World Energy Engineering Congress, AEE: Austin, TX, September 14–15, 2005.
10. Brown, M.; Key, G.; Meffert, W. *Understanding, ANSI/MSE 2000*; Distance Learning and Professional Education, Georgia Tech: Atlanta, GA, 2000.
11. Benson, K.; Key, G. C&A Floorcoverings' Strategic Approach to Energy Management. In *World Energy Engineering Congress*; AEE: Atlanta, GA, 2003; November.

Manufacturing Industry: Activity-Based Costing☆

J. Michael Fernandes
Mobile Platforms Group, Intel Corporation, Santa Clara, California, U.S.A.

Barney L. Capehart
Department of Industrial and Systems Engineering, University of Florida College of Engineering, Gainesville, Florida, U.S.A.

Lynne C. Capehart
Consultant and Technical Writing Specialist, Gainesville, Florida, U.S.A.

Abstract
Traditional accounting and product costing methods usually take energy costs, put them into the category of overhead, and then allocate these costs based on the product's labor usage rate. Often, this does not correctly assign the true energy cost to a given product or cost center. A more modern approach is Activity-based costing which identifies specific cost drivers and then allocates costs based on the details of the facility's equipment, operation, and product mix. This paper will apply the activity-based costing methodology to allocate energy costs in a manufacturing environment. Energy costs at Double Envelope Corporation, located in Gainesville, FL, are used as an example of this new method. To the authors' knowledge, this is the first paper to specifically apply the principles of Activity-based costing to detailed energy costing and energy management in an actual manufacturing company.

INTRODUCTION

In many manufacturing operations, fifty percent of a product's cost can come from material and about five to ten percent from labor. The rest is classified as overhead. Overhead costs generally include all indirect costs such as material handling, warehousing, maintenance, quality control, engineering, setup cost, machine depreciation, and energy.

Most company financial departments use the traditional accounting method for product costing. This method assigns overhead cost based on a single overhead allocation rate—usually directly proportional to the number of labor hours a product consumes during production. Because it assumes that all products require the same proportionate amount of indirect activities based on labor hours utilized, the traditional method distorts product costs and often gives management a poor picture of product mix profitability within complex multi-product companies. This method does not specifically trace the individual indirect activity costs to either the products or the cost centers that require them. It treats all products the same, even though they are very different and require different types and amounts of support.

There is an alternative costing method that does directly trace all overhead costs to the specific activities and products that generate them. This new method is called activity-based costing or simply ABC. An activity is defined in terms of a function within the company. Each activity defined must be homogeneous and quantifiable. Examples of activities are: material handling, warehousing, setup, maintenance, engineering, distribution or personnel. Each of these activities consumes resources, such as labor, material, machine depreciation, energy, and capital. All the resources or dollars necessary to perform that function are pooled together into a single activity cost pool for that function. Therefore, each activity has its own cost pool.

Activity-based costing allocates overhead to products via cost drivers, which are the bases used to make the cost assignments. The first-stage cost drivers—or resource cost drivers—allocate resource costs to the various activity cost pools. Examples of first-stage cost drivers are: electricity use in kilowatt and kilowatt-hour, fuel use in gallons or therms, facility space leased in square feet, and contract maintenance directly in dollars. Thus, the first-stage cost drivers allocate a percentage of a resource cost to each activity based on the percentage by which that activity uses each resource.

The second-stage cost drivers—or activity cost drivers—allocate costs from the activity cost pools to the various products or cost centers. Examples of second stage drivers are: number of transactions, number of setups, number of machine hours, number of inspections, number of work orders, or number of labor hours consumed. Thus,

☆ This entry originally appeared as "Allocation of Energy Costs in Manufacturing Using Activity-Based Costing" in *Energy Engineering,* Vol. 94, No. 4, 1997. Reprinted with permission from AEE/Fairmont Press.

Keywords: ABC method; ABC accounting; Energy accounting; Allocating energy costs; Energy cost analysis; Energy cost management; Energy cost control.

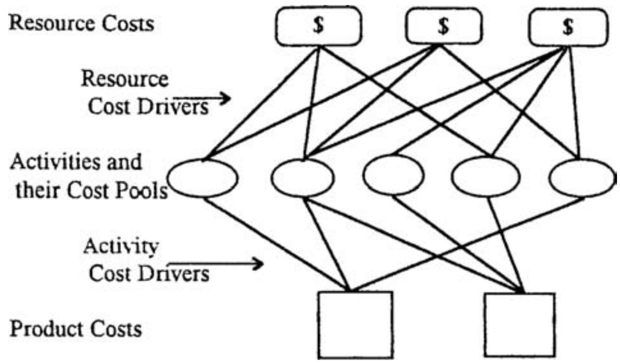


Fig. 1 Activity-based cost allocation.

the second stage cost drivers allocate a percentage of an activity cost pool to each product or cost center based on the percentage by which that product or cost center uses the activity. The sum of the allocations to a product or cost center is the total cost of the product or total cost allocated to the cost center. Second stage drivers assign costs to products or cost centers. The different types of costs assigned are unit, batch, product, organizational, and corporate costs. These second stage cost drivers determine the accuracy and complexity of the ABC system. Each second stage cost driver is unique to a given activity's cost pool. These drivers must be configured to match the activities they are associated with.

In summary, the first stage drivers—or resource cost drivers—allocate expenses, including energy costs, to activities and their cost pools. The second stage drivers—or activity cost drivers—allocate the activity cost pools to products or cost centers. When the activity cost pool contributions to each product are summed, they yield the actual product cost. Similarly, when the activity cost pool contributions to each cost center are summed, they yield the total cost allocated to that cost center. A graphical depiction of this allocation process is shown below in Fig. 1.

ENERGY COST ALLOCATION EXAMPLE

In general, activity-based costing allocates all resource costs in a company to activities or functions in the company. However, the ABC system can be used to determine the allocation of a specific resource cost throughout the company, such as energy costs. With an energy cost allocation system, management can determine which activities and products or cost centers are consuming the most energy and incurring the greatest energy cost. This information can make it easier to identify possible energy waste. Management can focus its efforts on the areas of large energy use and can evaluate them to see whether a redesign will reduce waste and increase energy efficiency. The ABC system also provides actual cost data that can be used to justify energy saving investments in specific areas or functions.

During the first part of 1996, a project was initiated to apply this ABC method for energy costing to an existing manufacturing company. Double Envelope Corporation in Gainesville, FL, agreed to supply data and to participate in this project. Double Envelope is a small business, which produces a variety of envelopes and folders for numerous commercial operations. One of the factors involved in selecting this company was that Double Envelope had started to record detailed machine hour data at each of its cost centers in 1995. This action made it possible to create an energy resource cost allocation based on ABC using energy cost drivers. Double Envelope makes a large number of specialized envelope products, and in order to simplify the effort needed to establish an ABC energy allocation system, it was decided to select six cost centers for detailed cost allocation rather than using the individual products.

Using the modified ABC system, we started by allocating the energy costs to activities or functions in the company. For Double Envelope, the only energy resource used was electricity, and the total energy cost including demand, electric energy, electric surcharge and taxes for 1995 was $183,253.14. The electric energy costs have been divided into five energy resource cost categories and treated as if they were individual resource costs. These five resource cost categories are lighting, cooling, motors, production machines, and miscellaneous use.

Our first task was to determine the actual costs for each of these five categories of equipment. This determination basically requires that a detailed energy audit and energy analysis be conducted for the facility. We counted lamps, determined the types of lamps (and ballasts if the lamps are fluorescent, high intensity discharge (HID) or low pressure sodium) used, and estimated the hours of use of the various lamps. We also collected similar information on wattages, efficiencies, and hours of use of the other equipment categories. Once this type of energy audit data is collected, it can be processed to determine the total energy costs in each of the equipment categories. However, as a practical matter, instead of finding the total cost of each equipment category and then determining how to split it up into the various activities, it is really easiest to locate and identify that part of the equipment which is utilized by the various activities and directly calculate the components of the activity cost pools.

Next we identified the various activities so we could calculate the activity cost pools. The four major activities at Double Envelope are warehousing, shipping and receiving, manufacturing, and general office. In a complete ABC analysis, energy use would be one of the first-stage cost drivers. Since we were studying only the electric energy use, we subdivided the energy cost driver into cost drivers for five groups of energy using equipment. The first-stage energy cost drivers for the Double Envelope analysis are the kilowatt-hours consumed by lighting, cooling, motors, production machines, and miscellaneous use for each of the four activities. For example, the lighting kilowatt-hour cost driver for manufacturing is 332,466 kWh of lighting per year. Once a cost driver is determined for a specific activity, the cost associated with that particular driver can be determined and allocated to that activity cost pool.

The total energy cost for any group of equipment such as lighting is calculated based on the total kilowatt-hour consumed. To be completely correct, the electric cost should be separated into demand (kW) and energy (kWh) costs, and those costs combined. However, the average electric cost was used in this project for simplicity. The total number of kilowatt-hour consumed at Double Envelope during 1995 was 2,229,600 kWh, and the electric bill was $183,253.14. Therefore, the average cost per kilowatt-hour is equal to $183,253.14 divided by 2,229,600 kWh, or $.0822/kWh. This figure of 8.22 cents/kWh was used in all the work that follows to calculate the costs of the five energy resource cost categories.

The lighting energy use and costs were calculated and separated into three plant areas, and were then allocated to the four activities using the kilowatt-hour consumed by each activity as the first-stage resource cost drivers. These lighting calculations are shown in Table 1 below. The manufacturing floor was lighted by 240 185-watt lamps. They were operated approximately 6240 h during the period. Adding in a 20% ballast consumption, the manufacturing floor activity used 332,466 kWh and

Table 1 Calculation of lighting and cooling energy resource costs

Cost per kilowatt hour

Total energy cost for 1995 = $183,253.14

Total kilowatt hour consumed during 1995 = 2,229,600

Cost per kilowatt hour = $183,253.14/2,229,600 = $.0822

Lighting

Area	Bulb type	Wattage	Number	Ballast usage	Total wattage	Usage time hours	Energy kilowatt hour	Cost 1995 ($)
Manufacturing floor	F96P617	185	240	0.2	53,280	6240	332,466	27,328.68
Warehouse	F96P167	185	36	0.2	7,992	6240	49,872	4,099.32
General offices	Fluorescent	40	176	0.2	8,448	2040	17,592	1,446.06

AC

Manufacturing floor and warehouse:

AC = Air conditioner size

H = Number of hours the air conditioner is in service per year

SEER = 9 Btu/Wh

CONV = Conversion factor of 3.412 Btu/Wh

AC = (12,000 Btu/h/tn) × (60 tn) = 720,000 Btu/h

H = 3000 h

Energy consumed = AC × H × 1/SEER × CONV

Energy consumed = 720,000 × 3000 × 1/9 × 3.412 = 818.88 MMBtu

Cost = 818.88 MMBtu × 1/3.412 Btu/Wh × 1 kW/1000 W × $.0822/kWh = $19,728

Warehouse = 0.2($19,728) = $3945.60 Manufacturing floor = 0.8($19,728) = $15,782.40

General offices:

AC = (12000 Btu/h/tn) × (20 tn) = 240,000 Btu/h

H = 2000 h

SEER = 7 Btu/Wh

Energy consumed = 240,000 × 2000 × 1/7 × 3.412 = 234 MMBtu/yr

Cost = 234 MMBtu × 1/3.412 Btu/Wh × 1 kW/1000 W × $.0822/kWh = $5,637.42

thus had a lighting cost of $27,328.68. The warehouse was lighted by 36 185-watt lamps for 6240 h, and used 49,872 kWh at a lighting cost of $4,099.32. The shipping and receiving area is located in the warehouse and takes up half of the warehouse floor space; therefore, half of the lighting cost was allocated to the warehouse activity and half to the shipping and receiving activity. The general office was lighted by 176 40-watt lamps for 2040 h at an activity cost of $1,446.06.

Next, the cooling or air conditioning cost was allocated to the three plant areas, and then to the four activities. These calculations are also shown in Table 1 at the end of the paper. The manufacturing floor and warehouse were cooled by four 15-ton air conditioning units with an average seasonal energy efficiency ratio (SEER) of 9 Btu/Wh. The units were estimated to run for a total of 3000 full-load equivalent hours over the period and consumed 240,000 kWh of energy. This resulted in a total cost of $19,728. It was estimated that the manufacturing floor activity consumed about eighty percent of the cooling energy. Therefore, $15,782.40 was allocated as manufacturing floor activity cooling cost. The remaining $3,945.60 was allocated as warehouse cooling cost. This cost was then allocated half to the warehouse activity and half to the shipping and receiving activity. The general office required 20 tons of cooling by smaller units with an average SEER of 7 Btu/Wh. They were estimated to operate approximately 2000 full-load equivalent hours over the year and consumed 68,582 kWh of energy. The resulting cost was $5,637.42 and was allocated as general office activity cooling cost.

Next, the costs of operating several machines with large motors was allocated to the four activities. The energy cost of the vacuum pumps, bailer machines and air compressor were calculated from the horsepower, efficiency, and usage of their motors. All calculations are shown in Table 2. The costs were calculated as follows: vacuum pumps—$27,210.00, one baler machine—$2403.00, the other baler—$2901.60, and the air compressor—$13,605.12. The costs of the two baler machines were allocated to the warehouse activity. The costs of the vacuum pump and the air compressor were allocated to the manufacturing floor activity. The miscellaneous energy cost category includes copy machines, computers and other small office items. It was estimated to be approximately 5% of the total energy cost which is $9,162.66. This cost was allocated to the general office activity. The remaining energy cost was $69,731.28, which was considered to be the cost of the only remaining

Table 2 Calculation of motors and production machines energy resource costs

Motors

EC = Energy consumed

HP = Horsepower

EFF = Estimated efficiency of motor

N = Number of motors

C = Conversion constant = 0.746 kW/hp

LF = Fraction of rated load at which the motor normally operates

H = Total hours of operation of the equipment over 1995

$EC = HP \times N \times 1/EFF \times LF \times H \times C$

$Cost = EC \times \$.0822/kWh$

Equipment	HP	N	EFF	LF	H	EC (kWh)	Cost ($)
Vacuum pump	40	2	0.9	0.8	6240	33,1026	27,210.00
Bailer machine	10	2	0.85	0.8	2082	29,238	2,403.00
Bailer machine (outside motor)	25	1	0.88	0.8	2082	35,298	2,901.60
Air compressor	40	1	0.9	0.8	6240	165,510	13,605.12

Miscellaneous

5% of Total Cost = 0.05 × $183,253.14 = $9162.66

Production Machines

The remaining cost is $69,731.28

All other costs have been identified

This remaining amount is assumed to be the production machine's energy cost

Liq–Mobil

Table 3 Energy costs allocated to activity cost pools through first-stage drivers

Activities	
Allocation	**Cost ($)**
(1) *Warehousing*	
50% of warehouse cooling	1,972.80
Bailer machine	2,403.00
Bailer machine	2,901.60
50% of warehouse lighting	2,049.66
Total	9,327.06
(2) *Shipping and receiving*	
50% of warehouse cooling	1,972.80
50% of warehouse lighting	2,049.66
Total	4,022.46
(3) *General offices*	
General office cooling	5,637.42
General office lighting	1,446.06
Miscellaneous energy costs	9,162.66
Total	16,246.14
(4) *Manufacturing floor*	
Manufacturing floor cooling	15,782.40
Manufacturing floor lighting	27,328.68
Vacuum pump	27,210.00
Air compressor	13,605.12
Production machines	69,731.28
Total	153,657.48

energy consuming activity; the production machine usage. This cost was allocated to the manufacturing floor activity.

Table 3 summarizes all of the energy costs allocated to the four activities. The total warehouse activity energy cost is $9327.06; the total shipping and receiving activity energy cost is $4022.46; the total general office activity energy cost is $16,246.14; and the total manufacturing floor activity energy cost is $153,657.48. Note that the manufacturing activity consumes over 80% of the total energy cost. This is fairly typical for many manufacturing companies.

Once the energy costs are allocated to the cost pools of each activity using the first-stage cost drivers, the second-stage cost drivers must be identified for each activity, and then costs from each activity cost pool are allocated to one of the six cost centers. The six cost centers at Double Envelope were wide range folders, wide range windows, web folding, specialty, printing and cutting. All calculations and allocations of energy costs from the four activities to the six cost centers via cost drivers are shown in Table 4. This specific process of allocating energy costs is shown in Fig. 2 below.

The manufacturing activity cost was allocated to the cost centers via the cost driver machine hours. Machine hours were used to correctly identify the usage of the manufacturing floor by each cost center. Warehousing is generally driven by product volume. The warehousing activity cost was be allocated to cost centers based on the total volume passing through each cost center. General office activity costs are organizational level costs which are always allocated based on volume. The shipping and receiving activity cost is generally allocated based on the number of orders or transactions processed by the department. This information was unavailable, so an estimate based on cost center volume was used. Since this activity cost was small, this estimate will not result in a significant error in the final results. If a full ABC system is developed for Double Envelope Corporation, a more accurate driver can be identified and used.

The final results of the energy cost allocations to each cost center are shown in Table 5, and are as follows: wide range folders—$49,500.12, wide range windows—$24,060.54, web folding—$37,177.02, specialty—$5150.94, printing—$32,433.12, and

Table 4 Data for second-stage cost drivers

Cost centers		
Cost center	**Total machine hours**	**Total volume (thousands)**
(A) Wide-range folders machines (101–105)	23,901.00	208,162.20
(B) Wide-range windows machines (131–140)	12,348.00	47,313.00
(C) Web folding	13,870.50	457,571.40
(D) Specialty machines (160–166)	2,682.00	7,272.00
(E) Printing	15,751.50	129,682.80
(F) Cutting	14,392.50	329,550.60
Total	82,945.50	1,179,552.00

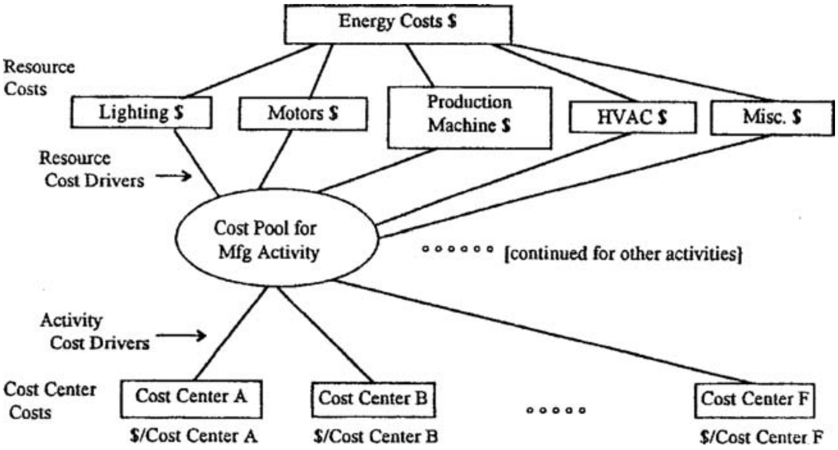


Fig. 2 Energy cost allocation at double envelope.

cutting—$34,931.04. This cost allocation data will give Double Envelope, an idea of which cost centers are consuming the most energy resources. By identifying a center's consumption, the company can decide if reductions should be made.

To show the impact of activity-based costing in this application, it is useful to compare the ABC allocation with the cost allocations that would be obtained with the traditional cost allocation of all overhead—including energy costs—using labor hours. What we have been calling machine hours is also the same measurement of labor hours for each cost center. Table 6 shows the results if the energy costs for each cost center are allocated proportionally to the labor hours utilized in each cost center. The comparison of these two approaches is shown below in Fig. 3.

The results in this example differ between ten to twenty percent for the ABC method vs the traditional method. This is a significant enough difference to show that the traditional costing method contains a cost distortion. In other examples, the cost distortion could be much greater. In the Double Envelope Corporation case, the large number of different products made it impossible to show individual product costing using the two methods. A study of a company with three to six individual products could illustrate results that show significant differences in final product costing.

For the energy manager of an organization, the ABC method provides an excellent opportunity to move the cost of energy from overhead to a line item in the cost of production, similar to the line item costs for labor and materials. Once the energy cost is known for each unit of production of a particular product, the energy manager will have a much easier time justifying new equipment and new processes that reduce energy costs. Since much of the new equipment and new processes would be expected to

Table 5 Allocation of activity costs to cost centers

Cost allocation

Cost center	Manufacturing floor ($)	General office ($)	Shipping/ receiving ($)	Ware housing ($)	Total ($)
A	44,276.88	2,867.04	709.98	1646.22	49,500.12
B	22,874.82	651.66	160.98	373.08	24,060.54
C	25,695.24	6,302.16	1560.72	3618.90	37,177.02
D	4,968.42	100.14	24.84	57.54	5,150.94
E	29,179.86	1,786.14	442.08	1025.04	32,433.12
F	26,662.26	4,538.94	1123.86	2605.98	34,931.04
Total	153,657.48	16,246.14	4022.46	9327.06	183,253.14

Allocation Equations

Manufacturing floor = (MH/82,945.5) × $153,657.48

General office = (Volume/1,179,552) × $16,246.14

Shipping/receiving = (Volume/1,179,552) × $4022.46

Warehousing = (Volume/1,179,552) × $9327.06

Liq–Mobil

Table 6 Traditional cost accounting allocation of energy costs to cost centers

Cost centers		
Cost center	**Total machine hours**	**Energy cost (thousands)**
(A) Wide-range folders machines (101–105)	23,901.00	52,804.95
(B) Wide-range windows machines (131–140)	12,348.00	27,280.68
(C) Web folding	13,870.50	30,644.37
(D) Specialty machines (160–166)	2,682.00	5,925.40
(E) Printing	15,751.50	34,800.10
(F) Cutting	14,392.50	31,797.64

increase production and also increase product quality, the data from ABC allows determination of the bottom line cost component of energy per unit of production. Thus, a large capital investment in new equipment and new processes which reduces the energy cost component of specific products can often be justified using the ABC method.

In the Double Envelope example, the number of individual products is so large that the ABC energy cost data has only been developed for the six major cost centers. However, even with this limited breakdown, some energy management decisions could be evaluated. For example, in the printing cost center (E), the ABC energy cost is $32,433.12 and the volume is 129,682,800 units. If a new printing press were under consideration by management, then the energy efficiency of the new press could be used to determine new energy costs per unit of production, and to help justify the additional cost of a more energy efficient machine.

CONCLUSION

Activity-based costing not only identifies the accurate cost of each product, but is a true decision making tool. Financial data should be accurate. Without correct information, it is impossible to make accurate decisions. Activity-based costing is a tool that can help companies become more profitable and understand the true functions and drivers of their costs. When ABC is used as a continuous improvement system, companies can identify the costs, implement quality programs, promote product and process redesign, reduce setup time, implement Just in Time (JIT), and get all of their costs under control. There are also many software packages available that can assist a company in implementing and managing an ABC system.

Although ABC systems are very useful, they cannot be used by all companies. Tracking all of the costs and activities can be very expensive. A company should only implement an ABC system when the information gained from the system is worth more than the cost of implementing and managing the system in the long run. As complex production management systems like MRP-2 and other software systems are further utilized in tracking production, the cost of data measurement will decrease. It is becoming more economical to implement a complex ABC system every day. As more systems are implemented, companies will become more efficient, profitable, and competitive in the world economy.

Cost Center	Traditional Cost	ABC Cost	%Diff from ABC
A	$52,805.23	$49,500.12	6.68%
B	$27,280.83	$24,060.54	13.38%
C	$30,643.43	$37,177.02	–17.57%
D	$5,925.00	$5,150.94	15.03%
E	$34,800.00	$32,433.12	7.30%
F	$31.796.00	$34,931.04	–8.97%

Fig. 3 Cost allocation comparison of results.

ACKNOWLEDGMENTS

The authors would like to acknowledge the cooperation and assistance of Mr. Steven Doyle and Double Envelope Corporation of Gainesville, Florida. Mr. Doyle was the production manager at Double Envelope Corporation during the course of this project, and assisted in providing the data on energy use and company operations. Double Envelope Corporation's assistance and interest in this project is greatly appreciated.

BIBLIOGRAPHY

1. Angelis, D. *The Effect of Activity-Based Costing On Traditional Operations Research Models*, Ph.D. dissertation, University of Florida, Department of Industrial and Systems Engineering, June, 1996.
2. Berliner, C.; James, B. *Cost Management for Today's Advanced Manufacturing*; The CAM-I Conceptual Design, Harvard Business School Press: Boston, MA, 1988.
3. Borden, J. Review of literature on activity-based costing. J. Cost Manag. Manuf. Ind. **1990**, *Spring*.

Liq–Mobil

4. Cooper, R. The rise of activity-based costing-part one: what is an activity-based cost system? J. Cost Manag. **1988**, *Summer*.
5. Cooper, R. The rise of activity-based costing-part two: when do I need an activity-based cost system? J. Cost Manag. **1988**, *Fall*.
6. Cooper, R. The rise of activity-based costing-part three: how many cost drivers do you need, and how do you select them? J. Cost Manag. **1989**, *Winter*.
7. Cooper, R. The rise of activity-based costing-part four: what do activity-based cost systems look like? J. Cost Manag. **1989**, *Spring*.
8. Hobdy, T.; Thomas, J.; Sharman, P. Activity-based management at AT&T. Manag. Account. **1994**, *April*.
9. Howell, R.A.; Soucy, S.R. Cost accounting in the new manufacturing environment Manag. Account. **1987**, *68*, 42–48.
10. Jayson, S. ABC is worth the investment. Manag. Account. **1994**, *April*.
11. Johnson, H.T. A blueprint for world-class management accounting. Manag. Account. **1988**, *June*.
12. Johnson, H.T.; Kaplan, R.S. The rise and fall of management accounting. Manag. Account. **1987**, *68*, 22–30.
13. McNair, C.J., Mosconi, W. *Meeting the Technology Challenge: Cost Accounting In a JIT Environment*; Inst of Management Accountants: Montrale, New Jersey, 1988.
14. Norkiewicz, A. Nine steps to implementing activity-based costing. Manag. Account. **1994**, *April*.
15. Navin, A.M. Applying overhead: how to find the right bases and rates. Manag. Account. **1992**, *March*.
16. O'Guin, M.C. *The Complete Guide to Activity-Based Costing*; Prentice Hall: New Jersey, 1991.
17. Roth, H.P. Are you distorting costs by violating ABC assumptions? Manag. Account. **1991**, *November*.
18. Ruhl, J.; Bailey, T.A. Activity-based costing for the total business. CPA J. **1994**, *February*. Smith, M. Managing your system. Manag. Account. **1994**, *April*.
19. Turney, P.B. Common cents: the ABC performance breakthrough. Cost Technol. **1992**.
20. Turney, P.B. Activity-based management. Manag. Account. **1992**, *January*.

Liq–Mobil

Measurement and Verification

Stephen A. Roosa
Energy Systems Group, Inc., Louisville, Kentucky, U.S.A.

Abstract
Measuring and verifying utility patterns is a complex, multifaceted process that often involves an engineering assessment of energy usage and costs. Measurement and verification (M&V) methodologies are used in performance-based contracts, project commissioning, and for certain project certifications. Thankfully, numerous technologies and methodologies are available to measure and document energy usage. Tools available to the energy engineer include a number of M&V guidelines and protocols that establish standards for primary M&V options, test and measurement approaches, and reporting requirements. The critical questions become "What really requires measurement?" and "What is the most appropriate approach to use for verification?" Using procedures identified in the guidelines and protocols, a M&V plan is developed to answer these questions and to serve as a guide as the process unfolds.

The purpose of this entry is to summarize and clarify the M&V process from the energy engineer's perspective. This M&V process typically involves five primary steps: (1) performing the pre-construction M&V assessment, (2) developing and implementing the M&V plan, (3) identifying the M&V project baseline, (4) providing a post-implementation report, and (5) providing periodic M&V reports. This entry clarifies this process and offers the energy engineer an overview of the planning aspects of M&V.

INTRODUCTION

Energy conservation projects are often justified by their financial strength as a result of the savings derived from their implementation. Prior to standardized Measurement and verification (M&V) procedures, little justification other than an engineer's calculations or a manufacturer's data sheet were used to justify the cost savings resulting from the improvements. Estimates of savings often vary based on calculation techniques, differences in application of technologies, site variables, and other differences in operating parameters creates a need to standardize procedures. In order to more fully justify longer-term capital improvements, the demand grew for M&V solutions that more precisely document project cash flows. In addition to savings from energy, there developed a need for M&V procedures that apply to water and sewer improvements, operations and maintenance savings, and savings from other opportunities.

A goal of the M&V process is to periodically validate both project savings and the performance aspects of projects. Criteria are established against which the success of meeting the performance requirements can be measured and assessed. Today, M&V is used to validate commissioning, meet requirements for Leadership in Energy and Engineering Design (LEED) certification, and assess performance contracts. When commissioning involves a guarantee of standard compliance, M&V is useful in providing guidance for baselines. For projects seeking LEED certification, M&V is recommended for lighting systems, wastewater applications, and credit under the energy and atmosphere category.[1] Perhaps the most widespread use of M&V is for performance-based contracts that guarantee savings. Performance contracts are negotiated agreements that guarantee energy and other project savings over a period of time, typically 5–20 years. Savings for other performance aspects of projects, such as operational and maintenance savings, may also be guaranteed. Often, the guarantees are provided by either the selected Energy Services Company (ESCO) or a third party. For performance contracts, the M&V process establishes standards to legitimize project savings and provides a means of comparing project cash flows to baseline conditions.

Measurement and verification services can be performed by the energy services provider, by the project owner, or by an indifferent third party. Measurement and verification provides performance assurance, meaning that there is a supportable rationale for the guarantees provided in the performance of the equipment installed. In order to bring credibility to the process, at least one member of the M&V team is typically a Certified Measurement and Verification Professional (CMVP). This certification program is co-sponsored by the Association of Energy Engineers and requires that candidates meet specific educational and professional experience criteria and pass a certification exam.

As a result of advancements in monitoring and measurement technologies, it is possible for energy engineering professionals to log and record almost every energy consumption aspect of the energy conservation measures they implement. Examples include the use of data

Keywords: Measurement and verification; Energy savings; Measuring energy savings; M&V; M&V options; M&V approaches.

loggers, infrared thermography, metering equipment, monitors to measure liquid and gaseous flows, heat transfer sensors, air balancing equipment, CO_2 measurement devices, and temperature sensors. Remote monitoring technologies are also available. With the widespread use of digital and wireless technologies, costs for monitoring equipment have been decreasing as the capabilities of monitoring technologies continue to improve.

In addition to improvements in measuring equipment, two primary M&V protocols have been developed in the United States. Early efforts included the North American Energy Measurement and Verification Protocol (NAEMVP, 1996) and the M&V Guideline for Federal Energy Projects (1996). Both remain in use, as they apply to many projects that were developed in the late 1990s. The M&V Guideline for Federal Energy Projects was updated in 2000, but it lacks applicability beyond U.S. government projects.[2] However, the North American Protocol was improved and evolved to international stature, becoming the current three-volume International Performance Measurement and Verification Protocol (IPMVP, April 2003). Volume I identifies methodologies and options for determining energy and water savings. Volume II concerns M&V concepts and practices for improving indoor air quality. Finally, Volume III deals with M&V approaches for assessing energy savings in new construction.[3] Each provides details of M&V options and guidelines for verifying energy savings. The IPMVP is the current state of the art and the most widely adopted standard.

AVAILABLE MEASUREMENT AND VERIFICATION OPTIONS

The theory of M&V in regard to assessments of resource usage over corresponding periods of time can be explained by the following equation:

$$\text{Change in Resource Use}_{(\text{adj})} = \sum \text{Post} - \text{Installation Usages} \pm \sum \text{Adjustments} - \sum \text{Baseline Usages}$$

Baseline usages represent estimates of "normal" usages prior to implementation of any cost savings improvements. Adjustments are changes in resource use that are not impacted by the improvements and are exceptional. Post-installation usage refers to resource consumption after the improvements have been performed. Using this formula, negative changes in resource use represent declines in adjusted usage, while positive changes represent increases in adjusted usage.

In order to quantify usages, specific measurement technologies and verification methodologies are selected. An understanding of the various M&V methodologies available is critical in order to propose the best M&V alternatives for projects. In performance contracting applications, the actual M&V approaches utilized are negotiated between the ESCOs and their customers. Since the IPMVP is the most widely used protocol, its standards will be used as an example. The four M&V options described in the IPMVP (April 2003) are summarized as follows:

Option A: Partially measured retrofit isolation.

Standardized engineering calculations (based on product lab testing by the manufacturer) are performed to predict savings using data from the manufacturer's factory testing and a site investigation. Onsite measurements are taken to quantify key energy-related variables. Noncontrollable conditions can be isolated and stipulated (e.g., stipulating hours of operation for lighting system improvements).

Option B: Retrofit isolation of end use, measured capacity, measured consumption.

Calculations are performed and retrofit savings are measured by using data from before-and-after site comparisons (e.g., infrared imaging for a window installation). Option B differs from Option A, as both consumption (usage) and capacity (output) are measured.

Option C: Whole meter or main meter approach.

Measurements occur at the main meter. Using available metered utility data or sub-metering, the project building(s) are assessed and compared to baselined energy usage.

Option D: Whole meter or main meter with calibrated simulation.

Option D is in many ways similar to Option C. However, an assessment using calibrated simulation (a computer analysis of all relevant variables) of the resultant savings from the installation of the energy measures is performed. Option D is often used for new construction, additions and major renovations.

To minimize the costs of measuring and monitoring the energy savings, a form of M&V Option A is often used. When higher levels of uncertainty or risk in quantifying savings are acceptable, Options A or B are often selected. Measurement and verification costs for these options can be mitigated somewhat by using statistically significant random samples when there is a large number of similar components involved (e.g., lighting fixture retrofits).

Uncertainty can be reduced when Options C or D are used. However, as more sophisticated procedures and technologies are employed to quantify energy savings, M&V costs tend to increase. Thus, decreases in uncertainty are purchased with increases in M&V costs. The added costs are due to the increased use of metering, sub-metering, digital equipment, and more complex monitoring systems (hardware) that aid in capturing information to quantify savings. Such equipment might include sensing devices such as temperature sensors, flow meters, and pressure switches; communications equipment; or programmable controllers. Options C or D may involve installation of remote monitoring and connections to existing digital energy management systems. More complex technologies simply require more data gathering, calibration, maintenance, and data management (software) capabilities.

THE MEASUREMENT AND VERIFICATION PROCESS

Implementing M&V strategies in energy performance contracts is a means of verifying the achievement of energy cost savings guaranteed in the contract over a period of time. Beyond satisfying the enabling legislation in states where M&V is required for certain types of projects, properly applied M&V can aid in accurately assessing project savings. Other benefits include allocating risks to appropriate parties, reducing uncertainties to acceptable levels, monitoring equipment performance, finding additional savings, improving operations and maintenance savings, verifying that cost savings guarantees are being met, and allowing procedures for future adjustments as needed. Primary steps in the M&V process typically include the following:

Step 1: Performing a Pre-Construction Measurement and Verification Assessment
Step 2: Developing the Measurement and Verification Plan
Step 3: Establishing a Measurement and Verification Baseline
Step 4: Providing a Post-Implementation Measurement and Verification Report
Step 5: Providing Periodic Assessment Reports

Each of these steps will be discussed in detail. For purposes of clarity, the five steps are described as sequential. Steps 1, 2, and 3 may occur sequentially or simultaneously during the beginning of the M&V project. Step 2 provides an action plan which defines subsequent activities. Steps 3, 4, and 5 require reporting to document results. Step 4 occurs shortly after project completion, while step 5 occurs periodically. Step 5 will use data to compare utility usage and costs to the baseline developed in Step 3.[4] Regardless of timing or sequence, the components of all five steps are critical for a successful M&V project.

STEP 1: PERFORMING THE PRE-CONSTUCTION ASSESSMENT

During project programming, energy and water saving measures are selected for implementation. In the view of many facility owners, the bundle of energy conservation measures (ECMs) developed for a performance contract can become complex. Energy conservation measures may include mechanical system improvements, controls, lighting upgrades, more efficient plumbing fixtures, and building envelope modifications. In performance contracts, bundling energy and water saving measures is often a means of matching specific financial goals to customer equipment needs. This creates the opportunity to combine short-term payback energy measures with longer-term payback measures. The end result is a mid-term payback project that matches the optimized financial term of the project. However, bundling measures can cause savings assessments to become more complex, particularly when multiple ECMs impact a single building. In such circumstances, the interactive effects on savings from multiple ECMs must be calculated. During the pre-construction M&V assessment, the physical site conditions that impact energy, water and sewage usage are identified and documented. This assessment provides an inventory of energy-consuming devices and equipment found on site. The inventory will include not only equipment that may be modified or replaced as part of the project but also other equipment whose energy usage has potential to necessitate an adjustment during future M&V reporting periods.

The pre-construction assessment is codified in a report that becomes a part of the initial project proposal. This report documents existing site conditions, inventories energy- and water-consuming equipment, and records operating conditions and parameters. The pre-construction assessment is often developed concurrently with the project proposal. The project proposal includes a listing of individual energy measures with identified costs and savings. When savings are discretely identified and separable, M&V for each of the measures can be performed on an itemized basis. However, due to the potential interactive effects among ECMs, the savings attributable to the measures may not be individually separable. Building owners are typically more concerned with total project savings than with savings from individual measures. As a result, savings from the selected measures are often aggregated during the M&V process. The aggregated savings are the total savings that accrue from the combined project.

STEP 2: DEVELOPING THE MEASUREMENT AND VERIFICATION PLAN

After assessing site conditions and identifying the measures to be implemented, the M&V Plan is developed. The M&V plan serves as the cornerstone of the energy services guarantee process and is critical to the success of the performance assessment. It is an agreement that is negotiated between the contractor and owner that will be included as part of the final contract. The purpose of the M&V plan is to identify and codify the procedures, methodologies, measurement devices, standards, and processes that will be used to effect the M&V of the program. As a result, the applicable M&V standard must be identified in the M&V plan. As a minimum, the components of the M&V plan typically include the following:

- Term of the M&V project and the term that applies to each savings component.
- List of applicable facilities and equipment.
- Lists of ECMs that require M&V and any other improvements that are not subject to measurement and verification.
- Project utility usage and cost baseline.
- Requirements for measurement of functional capacity (output).
- List of existing and proposed energy consuming equipment.
- Methodologies for calculating costs and savings.
- Listing of applicable standards.
- Product and equipment warranty verification requirements.
- Equipment maintenance requirements and criteria noting responsible parties.
- Schedule for post-installation report and post-installation site inspections.
- Schedule for required periodic M&V reports (typically reconciled quarterly, semi-annually, or annually) with descriptions of allowable adjustment criteria.
- Owner responsibilities and obligations.

A statement as to how excess savings are to be distributed and how shortfalls are handled is typically included. While a few projects use only one M&V methodology; more often, multiple options are used to assess the multitude of technologies used in a performance contract. Measurement and verification Option Application Matrixes are developed to identify the selected M&V options used for each measure at each site location. Typically, a two-way matrix lists the selected energy, water and maintenance savings measures on the *X*-axis and the site locations on the *Y*-axis. Intersecting locations on the matrix identify the applicable M&V option (A, B, C, or D) to be used.

Criteria for energy usage adjustments are also provided in the M&V plan. For example, if light levels in occupied areas are below current standards and need to be increased, an energy baseline adjustment may be allowed to compensate for the added electrical energy required to meet the current standard. If an increase in the quantity of outside air supplied to an occupied space is required as a result of mechanical improvements, energy use will likely increase due to the necessity of conditioning a larger volume of air. Adjustments can be made available to offset the increased energy use necessary for compliance. In such instances, the energy services provider may be allowed a reasonable credit for the increase in energy use needed to meet the new standards. In this way, newly applied standards can be incorporated into the architectural and engineering solutions without causing project savings to be penalized. Other noteworthy adjustments to baseline utility usage that are identified in the M&V plan include additions or deletions of equipment, changes in the plug load energy usage (e.g., soft drink machines or computers), changes in occupancy or occupied hours, and building expansions or additions.

STEP 3: ESTABLISHING THE MEASUREMENT AND VERIFICATION BASELINE

Implementing the M&V plan involves data gathering, compilation of data, and documenting the results in a series of reports. The M&V baseline report provides an energy use profile along with a utility usage baseline. This document records usage patterns prior to the implementation of any energy and water savings measures. The goal of developing a baseline is to clearly identify pre-installation utility usages so that future usages can be compared over time. Without a clearly quantified baseline, future assessments of saving may be invalidated.

In the graph entitled "Theory of M&V Baseline Development," actual energy use increases over time above the pre-installation level. Despite an increase in actual facility energy use, reduced energy use is identified by the M&V process as the difference between adjusted usage and actual energy use.

Developing a facility energy usage baseline for performance based projects also involves the following:

- Defining a time period (typically a modified 12 month baseyear) for the facility's historical baseline usage.
- Selecting a period for project implementation, during which energy usage may or may not evaluated.
- Defining a date for the beginning of the performance assessment period.

The identified baseyear may be the result of utility bill analysis or calculated based on observations and known events that resulted in energy and water usage or a period of time.

Information compiled for the energy usage baseline typically includes a list of existing utility meters, a history of data compiled from utility bills, interval demand billing data, utility rate sheets, and any site information that would assist in rationalizing energy use patterns. When whole building assessment options (options C or D) are employed, all building equipment and operational conditions require documentation and are included in the final baseline report.

STEP 4: PROVIDING A POST-IMPLEMENTATION REPORT

After the physical implementation of the project, a post-implementation report is developed and provided. This report documents the differences between what was initially intended to have been installed and what was actually installed. As a result, one purpose of this report is to identify and resolve discrepancies between the intent of the project scope of work and the actual implemented project scope. This requires a review of the final proposal and a comparison of scope with as-built conditions.

More than simply verifying installation of equipment and providing a project "punch list," the post-implementation report identifies all scope changes that impact project utility savings and costs and quantifies their impacts. Commissioning activities, equipment calibration, and performance deviations are documented. From this information, it is possible to reassess the established post-implementation utility usage accordingly. In addition, the post-installation report often provides an inventory of any primary energy consuming equipment in place that was not impacted by site improvements. Electrical loads, equipment sizes, existing fixture counts, plug load survey and facility operating parameters may also be updated if site conditions have changed after the pre-construction assessment was performed.

One example is the identification of differences between light fixtures proposed (counts and energy use) and those actually installed. If the project involved installation of light fixtures and retrofitting a building's lighting system, this report would identify all proposed fixture modifications (providing fixture types, quantities, and lamp wattages), indicate the fixtures that were actually installed, and identify any differences. Another example might be a change in chiller specification due to the installation of a more efficient chiller than was originally specified.

STEP 5: PROVIDING PERIODIC MEASUREMENT AND VERIFICATION REPORTS

Periodic reports are developed and provided as the M&V plan is implemented over time. The periodic reports compare energy use during the reporting cycle to baselined energy use, most typically on an annual basis. However, quarterly and semi-annual reporting periods are not uncommon. These reports tend to be comprehensive, utilizing software analysis tools to incorporate the influence of key variables.

The periodic reports will document the performance of the project as compared to the stated goals by quantitatively tabulating comparative savings and costs. Actual energy use (in units of fuel used) is compared to baseline energy use as defined in Step 4. Of concern to the ESCO and their customer is the relationship between the actual savings resulting from the project and the savings that were guaranteed. How the distribution of shortfalls is handled depends on how the performance contract distributes project risk. Typically, savings shortfalls are distributed among stakeholders as defined in the M&V Plan or the performance contract. In most cases, savings for performance based projects will meet or exceed projections. Current energy usage is compared to initial period baseline energy use in order to estimate the savings resulting from the project for the reporting period. Regardless, the periodic reports often reflect adjustments that are necessary and available due to changing and perhaps unforeseen conditions that may have impacted energy usage during the reporting cycle. Potential adjustments may be allowable as a result of weather conditions, changes in electrical demand, changes in use of activities, building additions, or installation of additional energy consuming equipment, if permitted by the M&V plan.

Procedures for adjustments are identified in the M&V plan. For any given project, additional adjustments may be provided and customized as the special circumstances of the project dictate. Adjustments can be managed with the use of spreadsheets or computer software programs such as FASER™ and METRIX™.

CONCLUSION

The purpose of this entry was to summarize and describe the M&V process from the energy engineer's perspective. It was determined that there are certain M&V guidelines and protocols that establish standards, several M&V options from which to choose, and variable reporting requirements that may be applicable. This entry described the M&V process as having five primary steps: (1) performing the pre-construction M&V assessment, (2) developing and implementing the M&V plan, (3) identifying the M&V project baseline, (4) providing a post-implementation report, and (5) providing periodic M&V reports.

The M&V plan is the cornerstone of the process, since it identifies the standards against which the resulting project savings are compared. The types of information needed to develop an M&V plan were identified. Certain components of the process, including selecting from the available M&V options, defining the baseline, designing an M&V plan and implementing the plan, were discussed in detail.

In addition, a number of reports were identified that document critical aspects of the process. These reports included a pre-construction assessment, the M&V baseline report, the post-implementation report, and the periodic energy usage reports. In their sum, the energy use baseline report, post-implementation report, and periodic reports provide the basis for evaluating the performance aspects of the implemented project.

In the future, it is likely that M&V will prove to be useful for a broader range of technical applications. As digital technologies become more widely used, it is likely that the cost of data accumulation and management will decline as the accuracy of measurements improves. In addition, M&V reporting standards are likely to be refined and become more standardized.

ACKNOWLEDGMENTS

This entry is a revised and expanded version of a previously published article entitled "Measurement and Verification for the Real World II," presented at the World Energy Engineering Congress, Austin, Texas, in September 2005. The author thanks the Association of Energy Engineers for permission to republish.

REFERENCES

1. U.S. Green, Building Council. *Green Building Rating System for New Construction and Major Renovations Version 2.1 Reference Guide*. May 2003. pgs. 73, 103, 173–181.
2. U.S. Department, of Energy and Office of Energy Efficiency and Renewable Energy. *M&V Guidelines: Measurement and Verification for Federal Energy Projects*, Version 2.2. DOE/GO-102000-0960. September 2000.
3. IPMVP New Construction Subcommittee. *International Performance Measurement & Verification Protocol*, Vols. 1, 2 & 3. April 2003. Available electronically by accessing http:/www.doe.gov/bridge.
4. Roosa, S.A., AEE-ETE Conference, Energy Efficiency, Energy Markets and Environmental Protection in the New Millennium, Sopron, Hungary, 15 June 2001, The Energy Engineer's Guide to Performance Contracting.

Measurements in Energy Management: Best Practices and Software Tools☆

Peter Barhydt
Arctic Combustion Ltd., Mississauga, Ontario, Canada

Mark Menezes
Emerson Process Management (Rosemount), Mississauga, Ontario, Canada

Abstract
This article will focus on considerations—often overlooked—which can have a significant impact on measurements used in energy management. Users will learn how to optimize the accuracy and repeatability of gas and steam flow measurements, minimize flowmeter permanent pressure loss (PPL), and optimize temperature measurement accuracy. Methodologies and/or software tools will be presented that allow the user to quantify the impact of their existing practice and any ROI from adopting the recommended practice.

IMPROVE FLOW MEASUREMENT

Accuracy and Repeatability

Accurate flow measurement is vital to prevent billing disputes in custody transfer applications. Repeatable gas flow rates help optimize combustion[1] and compressor efficiency.[2] Steam and water flow rates are used in the "second and third elements" of feed-water control, and help improve efficient control of heat transfer processes. For all of these fluids, accurate and repeatable internal allocation meters are needed to evaluate the effectiveness of initiatives for controlling energy usage, identify areas for improvement, or detect problems such as leaks. In plants with multiple boilers or compressors, flow measurements are used to determine unit efficiencies, helpful in deciding which units to base load and which to swing.

What should users do to improve the accuracy and repeatability of their flow measurements? Obviously, worn or damaged equipment should be replaced. However, even with new equipment, significant problems are commonly found when the flowmeters are not correctly installed and when gas and steam flows are not density compensated.

FIXING COMMON FLOWMETER INSTALLATION PROBLEMS

Not Enough Straight Pipe

Most flowmeter technologies require straight pipe up- and downstream of the flowmeter to ensure that the flow is well "conditioned" (Table 1). Unfortunately, standards defining these requirements have evolved over time, so many existing installations do not have enough straight pipe. This can cause severe flow errors which are not necessarily consistent ("bias"), as they can vary over the flowrate (Fig. 1). Inserting sufficient straight pipe may not be practical or cost-effective, and flow "conditioners" and "straighteners" increase installation cost and cause high pressure losses.

Newer flow technologies require less straight pipe than "traditional" devices such as orifice meters. For example, Coriolis meters do not require any straight pipe, though the meters themselves can be large and heavy. Magnetic and vortex flowmeters also typically require much less straight pipe than an orifice meter. Vortex meters are often preferred in applications with dirty fluids or wet steam, since they rarely suffer from erosion and newer designs are non-clogging. Unfortunately, many users have avoided vortex technology in the past due to its lack of flexibility when compared with an orifice plate. If the flow range used to size an orifice meter is incorrect, or if it changes over time, the user can usually accommodate the new flow range by re-ranging the pressure drop (DP) transmitter, and possibly replacing the plate. With a vortex meter, the user would not only need to change the entire body—even this is not too expensive, since presumably the old meter can be re-used in another application—but change the piping. The vortex meter requires straight up- and downstream piping, of the same dimension as the meter itself.

Newer reducer vortex meters avoid this problem (Fig. 2). With these devices, the supplier welds on reducing and expanding flanges, and in a flow lab calibrates out any impact on accuracy. To replace an existing meter with a reducer meter (or vice-versa), the user could simply replace the meter—no changes to piping.

One reason users have traditionally preferred DP-flowmeters to other technologies is that the DP-flowmeter can

☆ This entry originally appeared as "Energy Management Measurement Best Practices and Software Tools" in *Energy Engineering*, Vol. 102, No. 4, 2005. Reprinted with permission from AEE/Fairmont Press.

Keywords: Measurement; Steam flow; Efficiency; Pressure loss; Density compensation; Repeatability; Accuracy; Vortex.

Table 1 Required straight pipe (orifice meter, $B=0.6$), per ISO 5167

Required straight pipe diameters		
Upstream disturbance	**Upstream**	**Downstream**
Single elbow	18	7
Two elbows, same plane	26	7
Two elbows, out of plane	48	7

Source: From McGraw-Hill (see Ref. 3).

be easily verified. A user who is not confident that a DP-flowmeter is accurate can simply:

- Physically inspect the primary element (plate, Annubar, venturi, etc.) to ensure no damage.
- Verify the calibration constant for the primary element (beta ratio, etc.).
- Using a pressure source, verify that a simulated zero and full-span DP input into the secondary element (DP transmitter) provides the correct 4–20 mA output.

The same approach is available on newer vortex flowmeters:

- Physically inspect the vortex shedder bar for damage or coating.
- Verify the calibration constant for the shedder bar (*K*-factor).
- Using a hand-held communicator, verify that a simulated zero and full-span frequency input provides the correct 4–20 mA output.

Newest of all is the "conditioning plate," shown in Fig. 3. This device provides comparable accuracy and pressure loss to an orifice plate, but with only two diameters of straight pipe. Since the conditioning plate directly replaces the existing plate, reusing the same transmitter and taps, this is usually the most cost-effective approach to fixing the accuracy of an existing installation.

In new installations, eliminating the need for straight pipe usually allows the entire flowmeter to be located in an accessible location, minimizing costly and troublesome impulse lines.

Orifice Plate Misalignment

Most users periodically remove and inspect key orifice plates to ensure that they have not eroded. While worthwhile, this introduces the risk that the plate will not be correctly centered in the pipe when it is re-installed. Unless the user is using a mechanism which is inherently self-centering—such as the compact orifice shown in Fig. 3—centering errors of $\pm 1/4''$ are common. While not significant in large lines, they can cause 3%–4% flow error in a 2″ line.[4]

DENSITY COMPENSATION OF GAS AND STEAM

In steam and gas flow applications, the user needs mass flow—evidenced by the use of mass units such as pound per hour or kilogram per second, or "standard" units—standard cubic feet per minute (SCFM), normal cubic meters per hour (NCMH). By assuming that the density is fixed, the user can obtain mass flow by multiplying this density by the volumetric flowrate measured by, for example, the DP, vortex, ultrasonic, or turbine flowmeter. Unfortunately, density is never constant in even the simplest of applications. Fig. 4 shows a schematic of a system in which minimal pressure or temperature variation is expected—steam flow out of a new, pilot-operated regulator, followed by two elbows and a short length of clean, straight pipe.

As calculated[5] by the freeware P&T.xls spreadsheet (Appendix A), pressure variation in this simplest of applications can be up to 2.5 psi, leading to a non-repeatable flow error of 4%. Because the relationship between pressure loss and flowrate is exponential, if flowrate is halved, the DP will quarter; if it doubles, it will quadruple. As a result, the user can receive steam of widely varying density—and hence, mass flowrate—on a minute-by-minute basis.

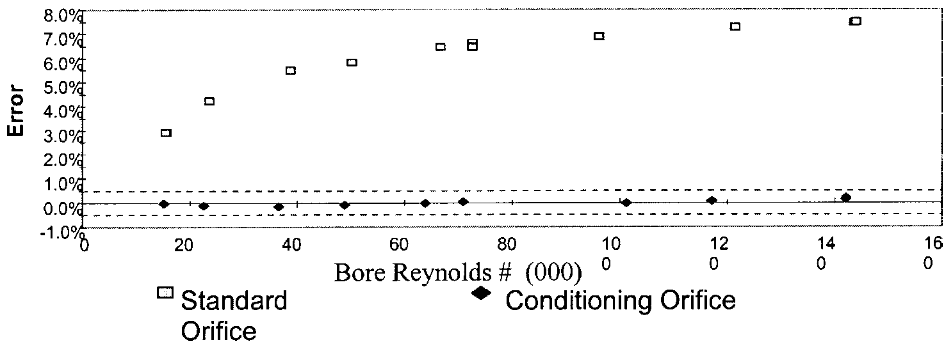


Fig. 1 Flow error from insufficient straight pipe (orifice meter).

Fig. 2 Vortex meters.

Temperature can also vary, even with saturated steam. As the steam flows, it loses pressure to friction, but—assuming well-insulated pipes—cannot lose enthalpy. This means that it must gain superheat, further reducing density. Temperature variation is even more significant with gases, since they are often sourced from ambient (outside) air, and gas pipes are rarely insulated.

To eliminate the impact of these variations, the "best practice" for any gas or steam flow application is to continuously measure pressure and temperature, and calculate density. Since no real gas is ideal, actual gas compressibility must be calculated using standards, such as AGA-3, ISO 5167, or steam tables. Fig. 5 shows a "multivariable" transmitter that measures differential pressure, line pressure, and temperature, and includes an integrated flow computer that calculates mass flow. The transmitter is shown integrated with an Annubar primary element.

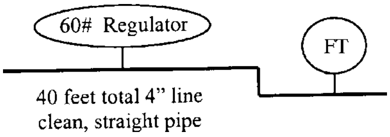


Fig. 4 Pipe schematic.

Minimize Flowmeter Permanent Pressure Loss

Fig. 6 shows a typical fluid pressure diagram for a differential pressure flowmeter. As the fluid enters the restriction, its pressure falls, creating a differential pressure (DP). This temporary DP (DP = 100 in H_2O) is measured by the differential pressure transmitter and related to the flowrate via Bernouilli's Law of conservation of momentum. Once the fluid exits the restriction, some of this pressure is recovered—the unrecovered pressure is the PPL. Permanent pressure loss for common industrial flowmeters can range from zero to greater than 20 psi.

In some applications, flowmeter PPL has no value. Obvious examples include flowmeters that are inline with mostly closed regulators or control valves. In the case of a mostly closed control valve, if the PPL through the flowmeter is reduced, the additional fluid pressure will simply be "scrubbed off" in the valve. In the case of a regulator, if PPL is too high the flowmeter can simply be moved from a location downstream of the regulator, to

Fig. 3 Conditioning plate—compact orifice paddle-style plate (with integral manifold).

Fig. 5 Multivariable™ transmitter with integral Annubar®.

Liq—Mobil

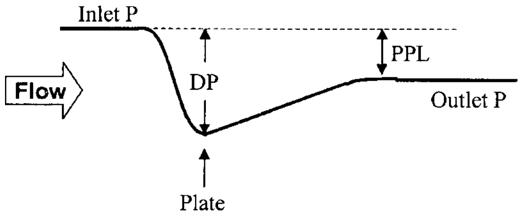


Fig. 6 Pressure loss through flowmeter.

upstream of the regulator. Applications where PPL has value fall into three general categories:

Increase in fluid pressure and/or flowrate = increased production rate.
Flowmeter inline with a variable speed/frequency drive = reduced energy cost.
New projects—smaller pumps and compressors = reduced capital cost.

Once the user identifies applications where reducing flowmeter PPL can provide value, the next step is to quantify the ROI achievable by replacing the existing or proposed technology with low-PPL technology. This calculation[6] is automated using the freeware Energy.xls spreadsheet shown in Appendix B.

Special attention should be paid to gas and steam flows with variable pressure. In most applications, demand for the gas or steam is based on mass flow. In other words, at a given production rate an application will require a certain NCMH of gas, or kilogram per second of steam. If inlet pressure is reduced, to obtain this required mass flow the volumetric flow must increase, increasing velocity and PPL. Consider an example—if inlet pressure is *halved*, PPL will *quadruple* through the meter, due to the squared relationship. While this may seem extreme, in many applications header pressure falls because demand—and flow rate—is high. To be safe, users should always calculate worst-case PPL at *minimum* line pressure as well as maximum mass flowrate.

COMPARING FLOW TECHNOLOGIES

In applications where the economic value of the average PPL is significant, or the worst-case PPL frequently limits production rate, the user should consider replacing a high-PPL technology with a low-PPL technology. The actual PPL of a specific flowmeter in a specific application is normally determined using meter-specific software. For standard devices such as orifice plates, this software is available from suppliers and third parties. As described above, the user must usually add the pressure loss from any required line-size pipe reduction/expansion to the software's calculated meter PPL.

In water applications where minimizing PPL is important, the recommended flow technology is usually the magnetic flowmeter, since it suffers from zero PPL. In gas and steam flow applications, the averaging pitot tube ("Annubar®") is usually the best technology. When compared with older round or curved designs, newer "T" designs provide much stronger signals, are easier to install, and much less likely to plug. As an insertion technology, they are also very cost-effective in larger line sizes, and are sometimes used for that reason in large water lines.

Optimize Temperature Measurement

Many industrial applications that consume energy do so to maintain a process at a certain temperature. If that temperature is measured inaccurately, the process can be under-heated (or cooled), possibly impacting quality or safety, or over-heated (or cooled), wasting energy.

IMPROVE ACCURACY OF TEMPERATURE MEASUREMENT

Resistance temperature detectors—RTDs—are usually specified in industrial applications that require high accuracy. Resistance temperature detectors rely on the principle that, for certain metals such as platinum, the resistance of the metal is linearly proportional to its temperature. Most users understand the need to compensate for resistance contributed by the lead wires, hence the widespread adoption of 3-wire, or even better, 4-wire RTDs. Less well understood is the need for "sensor matching."

When an RTD is connected to a transmitter or distributed control system (DCS) input, the user configures the transmitter to expect a certain sensor "type"—for example, "Platinum 100 ohm RTD." The transmitter or DCS uses this to convert measured resistance to temperature, using the IEC-751 standard. Unfortunately, no real sensor provides this ideal performance in the real world, and in fact the IEC-751 standard includes a tolerance which can exceed 3°F at higher process temperatures.[7] In applications where ignoring this "sensor interchangeability error" will reduce process safety or efficiency, the user needs to eliminate it by using "sensor matching." To accomplish this, the supplier of the RTD provides "Calendar van-Dusen" constants that define the "real-world" performance of a given individual sensor. These actual constants are then configured into the transmitter, eliminating the sensor interchangeability error and improving accuracy.

REDUCE COST OF TEMPERATURE MULTIPLEXING

Most energy management projects include a large number of new temperature monitoring points. Direct-wiring these points back to a central DCS or programmable logic controller (PLC), or using a large number of transmitters, can be cost-prohibitive. Instead, many users are using

Fig. 7 High density, standards-based temperature multiplexer.

high-density multiplexers. Newer multiplexer solutions (Fig. 7) use industry standards for communication, such as Foundation fieldbus and Modbus, avoiding proprietary devices with their risk of long-term obsolescence.

CONCLUSION

As users look to better manage energy flows, some often-overlooked best practices that should be adopted include:

- Ensure sufficient straight pipe or use conditioning orifice plates or reducer vortex meters.
- Ensure correct orifice plate centering, or use self-centering compact orifice meters.
- Density compensate gas and steam flows, using multivariable transmitters.
- Minimize flowmeter PPL by using averaging pitot tubes (Annubar).
- Optimize temperature accuracy by using Callendar-Van Dusen sensor matching.
- Reduce cost and avoid obsolescence by using standards-based temperature multiplexers.

APPENDIX A: P&T.XLS

The pressure and temperature (P&T) spreadsheet calculates:

- Pressure variation given user-entered flow conditions and piping schematic (including fittings) using Crane (1991).
- Flow error for DP and velocity/volumetric flowmeters due to this calculated pressure variation, plus any user-entered temperature variation.

Visit www.Rosemount.com/DP-Flow for further information.

APPENDIX B: PPL AND ENERGY COST CALCULATOR

The Energy.xls spreadsheet calculates PPL caused by various primary flow elements, and the associated annual energy cost. Visit www.Rosemount.com/DP-Flow for further information.

REFERENCES

1. Menezes, M. Calculating and optimizing repeatability of natural gas flow measurements. Pipeline Gas J. **2001**, *July*, 1.
2. Menezes, M. Improve compressor safety and efficiency with the right pressure transmitters. *Control Solutions*, Nov 2001; 1.
3. Miller, R.W. *Flow Measurement Engineering Handbook*; McGraw-Hill: Toronto, Ont., 1996.
4. Rosemount flow laboratory test results.
5. Crane Technical Paper 410Flow of Fluids through Valves, Fittings, and Pipe. 25th Printing, 1991.
6. Dieterich Standard Energy Savings Planner, Rosemount doc# DS-7105A.
7. Rosemount laboratory test results.

Mobile HVAC Systems: Fundamentals, Design, and Innovations

Alex V. Moultanovsky
Automotive Climate Control, Inc., Elkhart, Indiana, U.S.A.

Abstract
A vehicle heating, ventilating, and air conditioning system is not just desirable, but it is also necessary standard equipment for vehicles manufactured today. This system provides cab-crew comfort and is important for safety by ensuring demisting of the cab environment and defogging of windows in all kinds of weather. This entry is a sequel to the entry "Mobile HVAC Systems: Physics and Configuration" published in this Encyclopedia and it provides readers a short synopsis of mobile HVAC system psychrometric fundamentals, components design approaches, and components and system capacity calculations, as well as the future of mobile climate control systems.

DEFINITIONS AND NOMENCLATURE

For Definitions and Nomenclature please refer to entry "Mobile HVAC Systems: Physics and Configuration" published in this Encyclopedia.

INTRODUCTION

This article focuses on an overview of mobile HVAC system psychrometric fundamentals, and provides a short explanation of methods and approaches to system and components testing, capacity calculation, and design. At the end of this entry, prospective systems that take into account environmental impact and current no-idle regulations are presented.

PSYCHROMETRIC FUNDAMENTALS, AIR CONDITIONING AND HEATING CAPACITY CALCULATIONS, COILS DESIGN, AND MOBILE HVAC SYSTEM TESTING PROCEDURES

The amount of heat transfer Q (W, kW) that takes place as an air stream passes through a heating or cooling coil is the product of the mass flow rate of the air G (kg/s, sometimes kg/h) and the change in its heat content h (J/kg). During a cooling and dehumidifying process, the air stream undergoes changes in both sensible and latent heat contents. Sensible heat is the heat associated with the change in the air dry-bulb temperature. Latent heat is that amount released by the water vapor as it condenses. Sensible and latent heats are usually combined and expressed as total heat, or enthalpy.[1]

A change in enthalpy includes changes in both sensible and latent heats, but it does not account for the small amount of heat in the condensed water that has left the air stream. For some applications, however, the enthalpy of the condensed water is very small in comparison with the total enthalpy change, and as such, it is neglected for simplicity. When the air stream is heated or cooled without dehumidification, no latent heat transfer takes place. In this case, the heat transfer is said to be all sensible, or total heat transfer is sensible heat transfer.

The air-side heat transfer is balanced on the tube side of the coil, with an equal amount of heat being either absorbed or given up by the fluid flowing through the tubes. Again, the heat transferred is the product of the fluid flow rate G (kg/s or kg/h) and a corresponding change in fluid enthalpy h (J/kg). The fluid enthalpy change may be due to a temperature change only (for single-phase fluids such as water, glycol, oil, etc.), a phase change (evaporating or condensing refrigerants, steam, etc.), or a combination of both.

Psychrometric charts and tables usually present enthalpy, specific volume, and other moist-air properties on a dry-air basis rather than the actual air–water vapor mixture. This convention was adopted because it eliminates having to work with a varied air mass flow rate.

During a dehumidifying process, the total mass flow rate of the air–water vapor mixture changes as the vapor condenses out of the air stream. Because of this vapor loss, the mass of the mixture leaving the process is less than that which entered. The amount of dry air in the air stream remains constant throughout the process, however. Using air properties expressed in terms of this dry-air portion enables us to use a constant air mass flow rate, resulting in much simpler calculations. Then we have total heat transferred

$$Q_t(\text{W}) = G(\text{kg/s}) \times \Delta h(\text{J/kg}) \quad (1)$$

Keywords: Mobile heating, ventilating, and air conditioning (HVAC) system; Compressor; Condenser; Receiver drier; Expansion valve; Orifice tube; Accumulator; Evaporator; Capacity; Coil inverse design; Calorimeter.

and sensible heat transferred,

$$Q_s(\text{W}) = G(\text{kg/s}) \times C_{p.a}(\text{J/(kg}_a \cdot \text{K)}) \times \Delta T(\text{K}), \qquad (2)$$

where G, an air mass flow rate (kg/s) of dry air; Δh, an enthalpy change of the air (J/kg); $C_{p.a}$, a specific heat of the air (J/(kg_a K); and ΔT, the temperature change of the air.

Flow Rate Conversion to kg/s

For most coil heat transfer applications, flow rates are not given in kg/s (or kg/h). Instead, units are chosen that are consistent with those used in the application of fans, pumps, valves, and other components of the system. Air flow rates are generally given as cubic meters per hour $G_1(m^3/h)$ or per second $G_1(m^3/s)$. On the tube side, the units are usually in liters per hour or per second (L/h, L/s) for single-phase fluids. For evaporating or condensing refrigerants, the flow rates may be given as kg/s or expressed directly in terms of watts. For proper calculation, it is necessary to convert flow rate in m^3/s (m^3/h) to kg/s (kg/h) of dry air:

$$G(\text{kg}_a/\text{s}) = v(\text{m}^3/\text{kg}_a) \times G_1(\text{m}^3/\text{s}) \quad \text{or}$$
$$G(\text{kg}_a/\text{h}) = v(\text{m}^3/\text{kg}_a) \times G_1(\text{m}^3/\text{h}), \qquad (3)$$

where v = the specific volume of the air, m^3/kg of dry air.

Standard Air

The preceding formula uses the actual air volume flow rate. Most air conditioning equipment, however, is rated on the basis of a standard air flow. Standard air flow is a concept that was established to maintain uniformity in the testing, rating, and application of this equipment. Its use permits relatively simple calculation procedures for determining the performance of coils, fans, and other products. The coil performance ratings always are based on standard air flow rate at sea level (T_a = 20.7°C and P_{atm} = 760 mm Hg) only. By definition, standard air has a density of 1.201385 kg/m^3. At sea-level pressure (760 mm Hg) this density corresponds to that of dry air at a temperature of 20.7°C. The corresponding temperature for moist air is higher and depends on the actual moisture content. For practical purposes, 21°C is the generally accepted base temperature.

Conversion of Actual Air Flow to Standard Conditions

Because the performance curves are based on standard air conditions, air-flow rates must be expressed at these conditions as well before the curves are used.

For normal heating and air conditioning applications, the Temperature-Altitude chart may be used when the air flow is given as an actual flow rate at a temperature and/or altitude other than 21°C and sea-level pressure.

The enthalpy of air also varies with altitude. For cooling and dehumidifying applications at high altitudes, total capacity (watts) calculations should be based on the air enthalpies taken at the altitude in question. Heating and all-sensible cooling calculations are affected to a lesser extent, and conversion to standard conditions is usually the only correction taken.

Airside Heat Transfer (Capacity) Calculations

Usually, the process of mobile HVAC system design includes as a most important stage the evaluation of evaporator and heater coils capacities. There are various methods of this evaluation. Here, we will describe a few of them.

1. *Wet-bulb* (*dewpoint*) *and dry-bulb temperatures measurement method* (*using psychrometric chart*). Coil samples equipped with devices to measure dry-bulb and wet-bulb temperatures should be mounted on the inlet and outlet (inside the calorimeter enclosure) of the tested device. To obtain h_1 (enthalpy at the coil inlet), find the intersection of the dry-bulb temperature (from the sampling device) and wet-bulb (from the sampling device) or dewpoint temperature on a psychrometric chart. From this point, follow to the left along the diagonal line to the enthalpy scale. This new point is h_1.

To obtain h_2 (enthalpy at the coil outlet), find the intersection of the dry-bulb temperature (from the sampling device) and wet-bulb temperature (from the sampling device). From this point, follow to the left, again along the diagonal line to the enthalpy scale. This new point is h_2.

The Δh is the difference in enthalpy ($h_1 - h_2$). Multiply this number by $G(\text{kg}_a/\text{s})$ to obtain the air-side cooling capacity (power) of the tested evaporator Eq. 1. Although not exact, this formula provides reasonable accuracy for the range of application. More accurate calculations using inverse heat transfer methods will be presented below.

2. *Condensate collection method*. This method involves weighing the condensate collected off the core over a given time interval. To determine enthalpy h_1, the approach mentioned in the previous section will be applied. Locate where the dry-bulb temperature (from the sampling device) and the wet-bulb (from sampling device) or dewpoint temperature intersect on a psychrometric chart. Then follow that point over to the left with a straight edge to find the corresponding enthalpy value (h_1) at those conditions. Also, follow the intersection point over to the right to determine how many grains of moisture are going into the core (moisture in).

To determine h_2, it is necessary to collect all the water that has condensed off the evaporator over a given time interval (usually 15 min) at the same stable conditions that were used to find h_1. Next, weigh the condensate with an accurate scale, and convert the number to grains of moisture per kilogram of dry air (moisture condensed).

Now take the difference between moisture in and moisture condensed; this value is moisture out. Then locate the intersection of grains of moisture per kilogram of dry air (moisture out) and dry-bulb temperature (average air discharge temperature from the thermocouple grid on the core face) on the psychrometric chart. Follow this point over to the left to find the corresponding enthalpy (h_2).

The Δh is the difference in enthalpy ($h_1 - h_2$). Multiply this number by G to obtain the air-side cooling capacity (power) of the evaporator.

Tube-Side Heat Transfer Calculations

1. *Multiphase fluids.*

Total Capacity : $Q(\mathrm{W}) = G(\mathrm{kg/s}) \times \Delta h(\mathrm{J/kg})$.

Evaporator/Condenser Capacity Calculations

To obtain h_1 (evaporator/condenser inlet side), usually used the refrigerant temperature and pressure measured before expansion device (TXV or OT) for evaporator capacity calculations or at the inlet of the condenser (desuperheated gas) for condenser capacity calculations. By means of the refrigerant Saturation Properties–Temperature table (evaporator) or the refrigerant Superheated Vapor–Constant Pressure Tables (condenser), or any available software, and using measured pressure and temperature, it is easy to find the required enthalpy of refrigerant entering the evaporator/condenser.

To obtain h_2 (evaporator outlet, suction side or condenser outlet, subcooling side), use the refrigerant temperature and pressure measured right after evaporator/condenser outlet. By means of the refrigerant Saturation Properties–Temperature table or any available software, and using measured pressure and temperature, it is easy to find the required enthalpy of refrigerant exiting the evaporator/condenser.

The Δh is the difference in enthalpy ($h_1 - h_2$). Multiplying this number by G to obtain the refrigerant- (tube-) side cooling capacity (power) of the evaporator/condenser. Be sure to remain consistent with the units.

2. *Single-Phase Fluids*. The tube-side heat transfer rate must balance the air-side rate. The following formulas are used to determine the capacity by means of the fluid flow rate and the fluid temperature change.

Water Coils (Heaters)

Total Capacity (heat transfer rate):

$$Q(\mathrm{W}) = N(\mathrm{kg/s}) \times C_{\mathrm{p.f.}}(\mathrm{J/(kg \cdot K)}) \times \Delta T(\mathrm{K}), \quad (4)$$

where N, a fluid flow rate; $C_{\mathrm{p.f.}}$, a specific heat of fluid; and ΔT, a fluid temperature difference between the coil inlet and outlet.

Inverse Heat Transfer Method Approach to HVAC System Coils Design, Identifications, and Capacity Calculations

As already discussed, the main elements of vehicle climate-control systems include the evaporator and condenser coils and the heater core. These items are of primary interest in the area of heat transfer identification or Inverse Heat Transfer Problems (IHTP) because most of the time, the heat transfer for each heat transfer's stream in the climate-control system is estimated either from empirical equations or from heat balance equations. The estimation of heat transfer by inverse heat transfer methods first and foremost makes it possible to get precise results of heat transfer boundary conditions because these boundary conditions are identified on the basis of real temperature measurements, statistically taking into account the measurement errors and using the stochastic approach to the inverse problem's solution.[3,5,6] The most considerable advantage of the inverse methods over other approaches to the parameter identification is that the inverse methods do not require explicitly taking into consideration any of the numerous parameters and variables that affect heat transfer inside the HVAC module.

For an understanding of the complexity in evaluating the evaporator capacity or in other calculations of the evaporator's thermal parameters, let us list the variables that one takes into consideration. These are (1) the area of all evaporator plates, tubes, and fins; (2) temperature differences between the refrigerant and air through the evaporator; (3) thermal conductivity of the evaporator's material; (4) thickness of the plates and fins; and (5) different types of heat transfer, such as heat convection to the evaporator tubes by refrigerant flow, heat conduction through the evaporator's walls into the fins, and heat transferring by forced convection to the air stream flowing through the evaporator.

It is nearly impossible (or at best leads to lack of accuracy) to consider or calculate all these parameters, which are just part of the variables that affect the heat transfer in the evaporation process. Attempts to do so, either by the empirical equations or by the heat balance equations, would result in inaccurate or even wrong evaluation. There are many such calculations—i.e., Kurosawa and Noguchi.[2] The authors of this article calculated air-side heat transfer through the evaporator by means of a "contact factor," which had been determined from the empirical equation

$$\mathrm{CF} = 1 - \exp\left(-\frac{\alpha \times S_{\mathrm{e}}}{Q_{\mathrm{a}} \times C_{\mathrm{p}}}\right), \quad (5)$$

where α, the heat transfer coefficient between the air and the outside evaporator surface; S_{e}, the frontal area of the evaporator; G_a, the air-flow rate; and $C_{p.a.}$, the specific heat of air at constant pressure.

Even from this short example, the calculation of heat transfer evidently is inaccurate to begin with, because Eq. 5 is empirical and, thus, inexact; then the heat transfer coefficient, α, has been identified inaccurately from the imprecise heat balance equations. In addition to the huge number of variables that must be taken into account during the coil design process, as described earlier, one more point should be considered: packaging constraints.

It is obvious that the mobile HVAC system geometrical design and optimization or heat exchanger performance evaluation, based on approximate heat transfer calculations and rough evaluation of heat fluxes (or heat transfer coefficients) by means of simplified heat (energy) balance equations, is unacceptable.

The inverse method approach does not require explicitly taking into account any of the aforementioned variables because the measured temperatures already appear as a product of all parameters and variables that are responsible for the heat transfer process in the HVAC unit. In other words, these measured temperatures implicitly include all heat transfer influences in the HVAC thermal system.

With all the preceding, there is no doubt that the solution of the IHTP for the identification of boundary conditions and the device material's thermophysical characteristics is the most appropriate approach to getting precise estimates of desired thermal parameters. The concept, idea, and some results of the inverse approach (and utilizing the specifically created method of the Adaptive Iterative Filter), as well as the optimization of mobile HVAC heat exchangers' geometrical parameters, material selection, and capacity calculation based on this approach, have been presented in references.[3,5,6]

Evaporator Core Identification and Inverse Design

As regards the design of air conditioning systems, there is no need to say that it is very expensive and requires significant development work, even after the first prototypes have been built. To overview the complexity of the traditional approach to this design, let us just note that it requires the construction of a very complicated mathematical model of the evaporation process. This mathematical model of heat transfer between the evaporator and the air coming through it should include the following equations:

- Heat transfer from the air to the inner surface of the device
- Heat conduction through the evaporator's plates and fins
- Forced heat convection with air
- Heat exchange between the air and the device body inside the evaporator
- Heat transfer from the outer evaporator surface to the ambient air

To reduce the cost of design and development, eliminate unnecessary tests, and make the design of the evaporator core more precise, the idea of using inverse methods for identification of thermal parameters between heat exchanger and heat transfer media also has been implemented.[5,6] This approach allows reducing to a minimum the complicated calculation because it performs at the preliminary design stage and, as such, requires minimum time at a minimal cost.

It is apparent that the greatest factor that affects the evaporator performance (heat transfer and capacity) is the coil's dimensions. These dimensions, as already mentioned, are limited by the packaging constraints, which for vehicle applications are among the most important decisive factors in HVAC system design. The selection of the coil type depends in great part on space availability. The greater the space, the better the evaporator's overall heat transfer and capacity. On the other hand, the room available for system design relating to the evaporator sometimes appears to be too large for the required heat transfer or cooling capacity. It is clear that in this case, the selection of a smaller evaporator leads to savings, which in turn make the product more competitive. The concept of inverse approach to the optimization of heat exchanger size has been proposed in Moultanovsky 2002.[5] As is evident from previous reasoning, the evaporator's heat transfer and capacity are functions of a specific evaporator core. The coil's total heat transfer depends upon the air mass flow through the device and, specifically, through the device's frontal area. At the same air-flow rate, the heat transfer is defined by the surface area of the evaporator. Actually, the frontal area, as mentioned above, should be separated from the whole apparatus, because the dimensions of this area by and large determine the evaporator performance.

To summarize the influences of heat exchanger dimensions on the device's heat transfer, one can conclude that the greater the device's surface area, the better the whole apparatus performs, and vice versa. Very often the evaporator's required heat transfer can be satisfied by a coil with smaller frontal area than the available space. This statement immediately leads to the approach of the design optimization by means of inverse problem (the so-called inverse design).

An interesting approach to evaporator design optimization that shows considerable promise presented in Moultanovsky 2002 and 2001.[5,6] It was proposed there the evaporator material design and optimization of evaporator dimensions.

Optimization of the evaporator dimensions was based on the following considerations. The total heat transfer of the heat exchanger depends on the air mass flow coming through the device and its frontal area. So as mentioned above, at the same air-flow rate, the heat transfer is defined by the surface area of the evaporator. Based on this statement, Moultanovsky 2001[6] presented the construction of a nomogram that serves as an interconnection between the dimensions or, to be more specific, the size of the frontal area of the evaporator, air-device total heat transfer, and the evaporator's outer surface temperature. This nomogram enables the designer to choose the preliminary inverse selection (design optimization) of evaporator dimensions, satisfying the total heat transfer value.

As far as the frontal area of the heat exchanger is limited by the space available for the device, this area cannot be varied much to satisfy the requirements for heat transfer. That is why the way for HVAC system to significantly change the heat transfer between the evaporator and the air coming through it, is to produce the item from another material, thereby changing the heat transfer between the device and the surroundings. The material used in evaporator construction is of utmost importance because it greatly affects the thermal conductivity to the evaporation heat transfer process and, thus, the evaporator capacity. Estimation of heat transfer for the evaporator produced from alternative prospective material requires the evaporator's thermal system mathematical simulation. The latter is impossible without very accurate information about boundary conditions between the existing evaporator and the air. Based on the inverse problem approach to identification of these boundary conditions, Moultanovsky 2002 and 2001[5,6] proposed the construction of similar nomograms that serve as interconnections between the evaporator material (thermal conductivity), air-device heat transfer, and the evaporator's outer surface temperature (or air-flow rate). These nomograms enable the designer to make the preliminary inverse selection and inverse design of evaporator material. Basically, if given the minimal possible heat flux for the specified size of the evaporator, it is easy to find the evaporator's material, which provides the required heat transfer.

At the end of the evaporator discussion, it would be fruitful to compare the inverse problem approach of the evaporator capacity calculations with the above-mentioned calorimeter-based calculations. Let us take a quick look at traditional calculations:

1. The air-side heat transfer and capacity are obtained by means of measured air dry-bulb and wet-bulb temperatures, or dry-bulb and dewpoint temperatures, at the inlet and outlet of the evaporator. These measurements and psychrometric charts enable us to calculate the change of enthalpy of the air between the device inlet and outlet.
2. The desired capacity is a result of the multiplication of the air-mass flow rate and the enthalpy change.
3. Within this common approach, the heat of the condensed water that has left the air stream will be neglected for simplicity. To take into account the value of this heat, the humidity ratio has to be included in the calculation process. (The humidity ratio is a ratio of the mass of water vapor to the mass of dry air.) As a result, extra empirical calculations are required.

It is clear that the enthalpy changes and relative humidity values obtained by means of empirical psychrometric cannot be precise enough to satisfy the growing requirements for accuracy of total heat transfer and/or the capacity of the evaporator.

Capacity Calculations by Means of Inverse Heat Transfer Problem Approach

In fact, the heat transfer and capacity calculated using the IHTP approach make it possible to get the required precise results of desired heat transfer. The heat transfer identification procedure once again is based on accurate measurements of evaporator surface temperature that are already a product of the influence on the device of all three heats: sensible, latent, and condensed water. As a result, the coil's capacity calculations based on the outcomes of these measurements will include all kinds of heat that affect the performance of the cooling device.[5,6]

Heater Core Identification and Inverse Design

The heat transfer process in the heater core is less sophisticated than that process in evaporator core. The air side of the heater core is much more "controlling" than the coolant side because of the much lower air-side heat transfer coefficient (at least 1.5 times lower). This air-side coefficient greatly affects the transfer of heat from the coolant to the air; therefore, its accurate measurement or identification is of extreme importance. It should be noted once again that this coefficient of heat transfer or corresponding heat flux cannot be determined with sufficient accuracy because of the inaccuracy of traditional methods and, in fact, the incorrect approach to the determination. Following is an explanation of the traditional approach and the methods being used.

The different modes of heat transfer in the heater core consist of forced and natural convection, conduction, and radiation. The hot coolant transfers the heat by convection to the heater tube wall, where it is conducted through the wall into the fins. Then the heat is transferred by forced convection to the air stream flowing through the heater.

It is practically impossible to take into account all the complicated heat transfer in the system under study and to determine the overall heat transfer coefficient by using the heat balance equation between the coolant and air sides and/or by utilizing empirical equations that take into consideration the coolant physical properties and flow. That is why identification of the heat flux or heat transfer coefficient on the basis of IHTP methods with accurate temperature measurement on the surface of the heater core makes it possible to get precise results on the thermal parameters under study.[3,6] The flux identified in such a way is a function of a specific heater core itself. The obtained heat flux (or heat transfer coefficient) is used for the simulation of the heater thermal system following by the calculation of heater capacity and evaluation of the coolant side heater core performance. The heater capacity and performance data are applied to the heater core manufacturing process, as well as to the whole HVAC system performance evaluation. Other applications of the obtained boundary conditions are the same as explained above for the evaporator design, such as heater material inverse design and coil inverse design optimization.

New Approach to Mobile HVAC Components Evaluation

During the design phase of a mobile HVAC unit, many factors play a key role in the evaluation of system performance. The most decisive factors, however, are the actual performance of each component. These components include the evaporator, heater, and blower package. Any one of these components can have a dramatic impact on overall system performance. Traditional performance evaluation techniques do not allow for the all-around comparison of each component.

The special formulas and, respectively, special numerical values have been created (see Moultanovsky and Hermann 2001)[4] for each component comparison problem. The numerical value for component comparison is a Common Comparison Coefficient (CCC).

The traditional practice in evaluating blower and heat exchanger performance is to compare components based on air flow at a given backpressure or capacity at a given air flow. If the capacity of one of the coils is better than the capacity of another but pressure drop is worse, for example, the dilemma of which parameter is most significant for performance evaluation arises. Moreover, these two parameters are not the only factors that need to be integrated into the comparison. This list includes, but is not limited to, indoor and outdoor air temperature difference, both air and liquid-/refrigerant-side pressure drops, etc.

The same is true for blower comparisons. At least three parameters should be taken into consideration: airflow, rpm, and power consumption. Very often, they contradict one another, and as a result, the same dilemma arises.

In summary, each of the above-mentioned factors has an impact on performance independently. If any potential factor is left unchecked until final validation and testing, however, a significant failure of the HVAC system could go unchecked until the defects are correctable only by expensive and unnecessary redesign and tooling changes. Moultanovsky 2001[4] proposes a method of comparison through the use of weighting coefficients based on the emphasis of importance to the customer, designer, application, or manufacturer. This method of calculating performance will ensure that no factors or parameters have been overlooked. Common Comparison Coefficient calculations may have as many factors or parameters in the calculation as necessary to optimize the evaluation of overall performance. Overlooking these factors could become costly if they are not included in a performance evaluations. Common Comparison Coefficient is the numerical value that takes into account all compatible parameters produced during the testing of a product or a component.

The following calculation presents a comparison between component A and component B when component A is considered to be the baseline and its performance is accepted as 1 or 100%. It compares component A and component B using 3 different factors/parameters: Factor 1-X, Factor 2-Y, and Factor 3-Z.

It is clear that some parameters/factors are considered to have a positive impact on performance (such as capacity on heat exchangers or air flow on blower packages), whereas other parameters/factors will inversely affect performance (pressure drop through coils or power consumption on blowers). Let us assume that parameters X and Z are considered to have a positive impact on performance, whereas Y will be negative. The CCC can be calculated by the following formula (component A is considered to be the baseline):

$$\mathrm{CCC} = \frac{[K_1 \times (X_\mathrm{B}/X_\mathrm{A}) + K_2 \times (Y_\mathrm{A}/Y_\mathrm{B}) + K_3 \times (\mathrm{Z_B}/\mathrm{Z_A})]}{3}, \quad (6)$$

where K_1, K_2, and K_3 are the weight coefficients for each parameter according to importance. For more details on using this approach, see Moultanovsky and Herrmann 2001.[4]

MOBILE HVAC SYSTEM INNOVATIONS AND THE FUTURE

The next generation of air conditioning and heating systems is particularly important to the commercial success of electric, hybrid, fuel-cell, and other low-emission vehicles, which can capture market share only if they are equipped with good-performance, highly energy-efficient, and reliable cooling and heating systems.[7] Another, very important issue that significantly influences the future of mobile HVAC systems is a vehicle's anti-idling rules and regulations. These

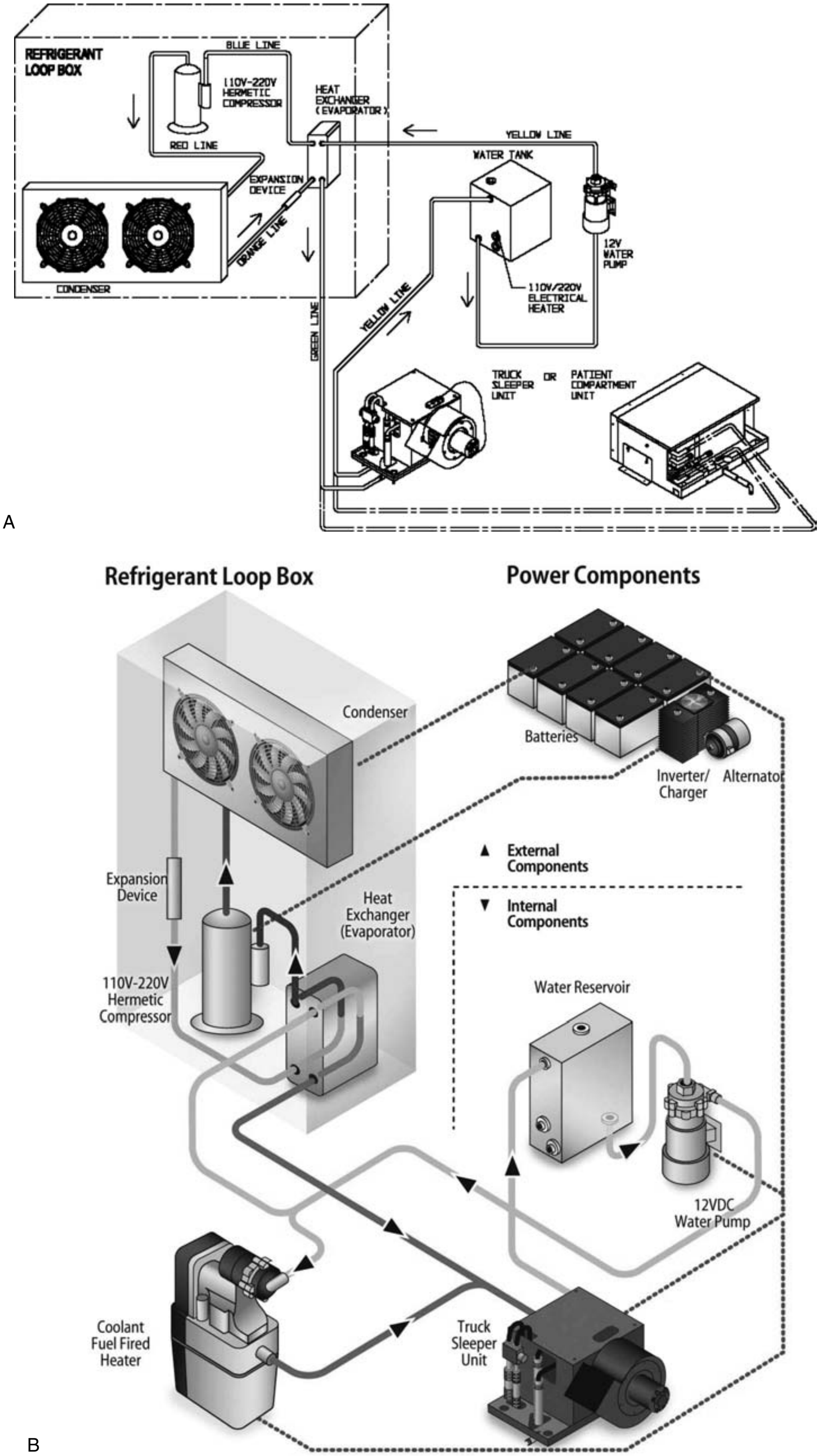


Fig. 1 Electrified, self-contained, secondary loop, hermetically sealed mobile HVAC system. (A) Schematic with electrical heating elements. (B) Complete system diagram with coolant fuel heater, internal and external components, and power package.
Source: From Courtesy of ACC Climate Control, Inc.

Fig. 2 Air-to-air fuel-fired heater.
Source: From Courtesy of ACC Climate Control, Inc.

regulations have an effect on many categories of vehicles, including trucks, motor coaches, transit buses, and shuttle and school buses. An average truck idles approximately 2400 h/yr, burning about 2400 gal of fuel during that period. Assuming that 1 gal of fuel costs $3, the cost of 1 truck idling its engine is about $7200/yr. Total idling time for all Class 8 trucks in the United States is about 1 billion h/yr, which accordingly results in 1 billion gal of diesel/yr or $3 billion. If the trucks are not idling, the engine regular service intervals will be significantly increased, and engine oil breakdown will be reduced. Major savings will also be realized as the time between scheduled major preventive-maintenance tear-downs will be increased. Last but not least is a pollution and greenhouse effect.

Strategies for reducing greenhouse-gas emissions and decreasing or eliminating idling is an electrical HVAC systems and, where feasible, shifting to hermetically sealed design.

Today's automotive AC systems are designed to use less refrigerant. To protect environment the US Environmental Protection Agency encourages all owners of the vehicles manufactured before 1995 that used HVAC systems with CFC-12 refrigerant to retrofit these systems to HFC-134a or other approved alternative refrigerants. New refrigerants that are being considered for future mobile AC systems are CO_2 (R744) and slightly flammable HFC-152a (to use with secondary loop systems).[7]

A very promising approach to mobile HVAC systems that provides a complete solution for the entire industry is a secondary loop system. This methodology makes it possible to build a self-contained, engine-driven AC system as well as an electrified, hermetically sealed system that complies with all no-idle regulations. An excellent example of such a system has been brought to the market by ACC Climate Control, Inc., Elkhart, Indiana. Fig. 1 presents a schematic (A) and complete diagram (B) of this secondary-loop, electrified system.

This revolutionary (for mobile HVAC systems) technology—a secondary-loop, hermetic system—includes 1 box containing a condenser coil, a hermetic electrical compressor, and a heat exchanger. Another part of this system consists an antifreeze reservoir with electrical water heaters (Fig. 1A) and an electrical water pump. Alternative source of the heat is a coolant fuel fired heater (Fig. 1B). Figure 1B represents the complete no-idle mobile HVAC system diagram with internal components (installed inside the vehicle), external parts (mounted on the frame, outside of very expensive inside compartment) as well as power package includes battery pack, inverter/charger (allows to use shore power when its available), and alternator. The beauty of this system is the installation connection process. When the components are installed on the vehicle and wiring is complete, heater hoses need to be spliced into the heater lines running from the engine to the auxiliary heater coil, using vacuum or electric water-flow valves. This system allows installing at low cost an extra heat exchanger (or heat exchangers) in the coolant line if additional air conditioning or heating points are required. This will eliminate the necessity for a less-efficient air ducting system. The system can be powered from a 110-VAC or 220-V power source, either from shore power or an auxiliary power unit on the vehicle, or from a 12/24/48-VDC battery package (as mentioned above, last system is designed for dual power: battery package and shore power). When the engine is shut down, the system operates at the same comfort level as when the vehicle (truck, bus, or specialty/emergency vehicle) engine was running. The secondary-loop approach eliminates refrigerant leaks and enhances high reliability. The system is environmentally friendly and compliant with all no-idle laws and regulations.

A new approach to satisfying vehicle heating requirements is the fuel-fired heaters brought to the mobile market by Espar, ACC Climate Control, and Webasto (Fig. 2).

These diesel fuel-powered heaters can serve as primary auxiliary heaters and ventilators while driving or as heaters without the engine idling for extended or overnight stops. The heaters are compact, lightweight, and quiet when operating. The heaters' efficient operation reduces both fuel costs and air pollution, and they comply with all no-idle regulations.

All future mobile HVAC systems, as well as systems under development, integrate cooling, heating, defrosting, demisting/defogging, air filtering, and humidity control. They increase driver alertness and visibility (demisting/defogging windows), as well as passengers' security. The success of air conditioning systems requires customer acceptance (cooling performance and reliability), operational and service safety, environmental performance, and serviceability. New mobile HVAC systems must anticipate future industry technology—such as higher engine efficiency (less waste heat and electrification) and electric, hybrid, fuel-cell, and low-emission vehicles—while being high-efficiency and satisfactory-performance devices. All vehicles benefit from efficient cooling and heating.[7]

ACKNOWLEDGMENTS

The author wishes to acknowledge the support of ACC Climate Control, Inc., which allows him to use materials and documents that are the property of ACC Climate Control.

REFERENCES

1. Hewitt, G.F.; Shires, G.L.; Polezhaev, Y.V. *International Encyclopedia of Heat and 20 Mass Transfer*; CRC Press: New York, 1996.
2. Kurosawa, I.; Noguchi, I. Development on a high efficiency drawn cup type evaporator core. In *Automotive Climate Control Systems, Selected Papers Through 1990*; Anglin, P.E., Ed.; Society of Automotive Engineers, Inc.: Warrendale, PA, 1991; 235–240.
3. Moultanovsky, A.V. Automotive Heater Core Boundary Conditions Identification, Proceedings of the ASME Heat Transfer Division-1998. ASME International Mechanical Engineering Congress and Exposition, American Society of Mechanical Engineering: New York, 1998; HTD-361; 5, 161–173.
4. Moultanovsky, A.V.; Herrmann, B. New approach to climate control system evaluation. J. Pass. Car.: Mech. Syst. **2001**, *6* (110), 592–598.
5. Moultanovsky, A.V. Mobile HVAC system evaporator optimization and cooling capacity estimation by means of inverse problem solution. Inverse Probl. Eng. **2002**, *10* (1), 1–18.
6. Moultanovsky, A.V. Design of the units and optimization of the components of mobile climate control system by means of identification of thermal parameters, Proceedings of the 4th International Conference on Inverse Problems: Identification, Design, and Control. The Russian Society of Inverse Problems in Engineering, Moscow Aviation Institute, Moscow State University: Moscow, Russia; 2001.
7. U.S. Environmental Protection Agency. Professional A/C Service and Repair. EPA US, MACS, and SAE International, 2003.

Mobile HVAC Systems: Physics and Configuration

Alex V. Moultanovsky
Automotive Climate Control, Inc., Elkhart, Indiana, U.S.A.

Abstract
Vehicle heating, ventilating, and air-conditioning (HVAC) systems are necessary standard equipment in today's manufactured vehicles. These systems not only provide comfort to the cab crew, but also are an important safety feature that ensures a demisted cab environment and that can defog the windows in all kinds of weather. This entry provides the reader a short synopsis of mobile HVAC systems, including each component's operation as well as the configurations of current and future systems.

NOMENCLATURE

Effective temperature a combination of relative humidity and temperature, which can indicate the level of comfort perceived by the human body.

Operator/passenger/patient (further—operator) enclosure part of the machine that surrounds the operator, preventing the free passage of external air, dust, or other substances into the area around the operator. This surround shall comprise components such as glass, roof, and floor.

Operator environment the space surrounding the operator, as defined by temperature and velocity measurement.

Air-conditioning (AC) system any system that lowers the effective temperature of the air within the operator enclosure.

Heating system any system that raises the effective temperature of the air within the operator environment.

Ventilate air change for comfort in the area around the operator in an operator enclosure.

Ventilating system any system that provides fresh air to and maintains air circulation within the operator enclosure.

Outside air/fresh air the unconditioned air entering the operator enclosure.

Recirculated air the air within the operator enclosure that passes through the AC system.

Recirc mode the HVAC unit operational mode uses 100% recirculated air.

Cooling decrease of the temperature of the air inside the operator enclosure.

Heating increase of the temperature of the air inside the operator enclosure.

Keywords: Mobile heating, ventilating, and air-conditioning system; Compressor; Condenser; Receiver dryer; Expansion valve; Orifice tube; Accumulator; Evaporator; Capacity; Calorimeter.

Pressurization pressure differential between the static pressure inside and outside the operator enclosure.

Pressurization system means used to pressurize the operator enclosure, including any components that influence the performance of the system.

Filtration removal of dirt and dust particles from the air forced or drawn into the operator enclosure by mechanical means.

Filtration system means of removing dirt and dust from the air entering the operator enclosure.

Filter efficiency measure of the capability of the air filter element to remove particulate matter.

Solar heating heating factor from the sun to be considered in determining air-circulation and cooling requirements necessary to maintain a comfortable temperature inside the operator enclosure. Solar heating must be considered carefully when designing the system. Solar radiant energy causes a heating effect to be considered in determining cooling capacity and insulating requirements. Also, direct radiation on the operator can cause discomfort not easily overcome by the system.

Solar radiant energy process by which solar heating is generated.

Ambient temperature the air temperature at which the equipment will normally operate is one of the most significant factors to be considered in determining the capacity required for proper heating or cooling of the operator enclosure.

Mechanical load heat transferred to the operator enclosure from the engine, transmission, hydraulics, etc. This load must be considered when locating system components, and it may require special insulation.

Dirt and dust must be considered when designing the filtration system and when locating components, particularly those outside the operator enclosure.

Global factors ambient conditions of the various markets that the vehicle will operate in. (Compare North America with Europe for ambient loads.)

Heat transfer transfer of heat from liquid to air; this heat is directly proportional to the difference between the temperatures of the liquid and air entering the transfer system for a given rate of liquid flow and airflow measured in kg/s or kg/hr and that heat removed from liquid is equal to the heat given to air.

Heating/defrosting system the system used to heat the cab interior and to clear the windshield, including all ducting, fans, and heat-exchanger equipment. This system is the means intended to defrost the windshield and specified portions of the right and left windows.

Defrosting melting frost on the inside of the glass, or on the test coating on the outside surface of the glass, with the defroster/demister system.

Defrosted area the area of the windshield and the right and left windows composed of dry, cleared surface and melted or partially melted (wet) test coating, excluding that area of the windshield covered with dry test coating.

Coolant liquid used for heat transfer, composed of 50% ethylene glycol and 50% water or other liquids specified by the vehicle manufacturer for use in the heat transfer system.

Heat-exchanger system provides heating and windshield defrosting and defogging capabilities in a vehicle. The system consists of an integral assembly or assemblies, with a core assembly or assemblies, blower(s), fan(s), and necessary duct systems and controls to provide heating, defrosting, and defogging functions.

Heat exchanger core assembly in general, the core consists of a liquid-to-air heat transfer surface(s), liquid inlet, and discharge tubes or pipes.

Evaporator assembly a coil (heat transfer surface) that forces air over the heat transfer surface into the cab and the complete enclosure (AC unit) to be furnished for the installation.

Heat exchanger/defroster duct system passages that conduct inlet and discharge air throughout the heat exchanger system. The discharge outlet louvers are part of the system.

Air delivery rate the actual rate of airflow (standard air volume flow rate as specified in "System Components and Duties") for wet coil conditions.

Cooling capacity (air side) the amount of heat absorbed from the air flowing through the evaporator (W and kW).

Cooling capacity (refrigerant side) the amount of heat absorbed by refrigerant flowing through the evaporator tubes (W and kW).

Dry-bulb temperature air temperature in °C (K) as read from a standard thermometer or other appropriate temperature-measuring device.

Wet-bulb temperature air temperature in °C (K), essentially equal to that read from a wet-bulb thermometer or one whose sensing bulb is covered by a water-wetted wick located in the moving ai stream.

Relative humidity ratio of the amount of moisture (in mol-fraction) present in the air to the maximum amount the air can hold at the same temperature and pressure.

Humidity ratio ratio of the mass of water vapor to the mass of dry air (kg_w/kg_a).

Dew point (saturation) temperature air temperature in °C (K) at which moisture begins to condense out as the air is cooled at constant pressure.

Specific volume ratio of the total volume of moist air to the mass of dry air (m^3/kg_a).

Specific heat the amount of heat required to raise the temperature of a kilogram of a substance (air, liquid, etc.) by 1 K. Specific heat is a function of temperature. $C_{p.a}=$ Specific heat of the air in $J/(kg_a \cdot K)$.

Specific gravity of liquids ratio of the mass of liquid to the total volume of liquid at a certain temperature (kg/l).

Enthalpy (sometimes total heat) a convenient energy concept defined from the properties of a system. Enthalpy describes how much heat a substance contains, determined from a predetermined base or starting point. Often, enthalpy refers to the total heat (not very accurate statement) content (sensible and latent) present in the air (J/kg_a) equal to the sum of the individual partial enthalpies of the dry air and water vapor.

Sensible heat the amount of heat associated with a change in the dry-bulb temperature of the air.

Latent heat the amount of heat required to change the state of a substance. Specifically, it is the heat released as water vapor condenses out of moist air and also the heat associated with the phase change of certain heat transfer fluids (volatile refrigerants, steam, etc.).

Entropy a microphysical property of thermodynamic systems. It is completely transferred from one system to other during a reversible process, whereas it always increases during an irreversible process in a closed system.

Discharge line high-pressure (inlet) line that carries liquid refrigerant to the expansion device.

Suction line low-pressure (outlet) line that carries evaporated (gaseous) refrigerant to the compressor.

Subcooling the degrees of temperature below saturation (based on the inlet pressure of the expansion device) of the liquid refrigerant (°C or K).

Superheat the degrees of temperature above the saturation temperature (based on the outlet pressure of the evaporator) of the vaporized refrigerant (°C or K).

Expansion device a valve or fixed orifice in the refrigerant circuit with the purpose of metering refrigerant into the evaporator, inducing a large pressure drop and causing a change of state.

Compressor a device that pumps low-pressure refrigerant vapor out of the evaporator by suction, raises its pressure, and then pumps it under high pressure into the condenser.

Condenser a device that removes heat from the entering high-pressure, high-temperature, desuperheated refrigerant vapor, changing it to a high-pressure, high-temperature liquid.

Evaporator a device that removes unwanted heat from the air by boiling liquid refrigerant in the evaporator coil.

Psychrometric chart a graphical presentation of moist-air properties.

Sleeper cab the occupant space behind the truck cab intended to be used as living space during travel or while the vehicle is parked.

Calorimeter a test fixture including a specially built air tunnel that directs all the air from the discharge face of the evaporator, heater, or condenser (or complete HVAC unit). The sampling device is set up at the end of the tunnel. The calorimeter is equipped with an AC and heating loop to simulate the HVAC system operations. Most of heat transfer parameters denominations correspond to reference.[1]

INTRODUCTION

An HVAC system is standard equipment for most of today's manufactured vehicles. The mobile HVAC system's main duty is to provide driver and passenger temperature comfort and to provide the required cab dehumidification during hot, high-humidity weather without failure due to device cooling capability, all while still allowing safe vehicle operation. It must cool down the vehicle enclosure within a short time, often within 30 min, at ambient temperatures up to 45°C–50°C. The dehumidification feature in great part determines the required system's (evaporator's) cooling capacity. The heating part of a mobile HVAC system should provide acceptable defrost, demist, and defog performance, yet the heater must make driver and passengers comfortable within a relatively short period after a cold startup in temperatures like −30°C and lower (as in some cold areas of the United States or the world). The system must be quiet, and the controls must be easy to understand and operate.

This entry focuses on an overview of mobile HVAC system, including descriptions of each component. Also, in the entry "Mobile HVAC Systems: Fundamentals, Design and Innovations" of this Encyclopedia will be provided the explanation of methods and approaches to system and components testing, capacity calculation, and design. At the end of that entry, prospective systems that take into account environmental impact and current no-idle regulations are presented.

HISTORY

The thermodynamic basis for modern AC began when the French military engineer Sadi Carnot (1796–1832) formulated the basic principles of the reversed Carnot Cycle. The aim of all clockwise-operating cycle processes is to produce work by transferring heat from a high-temperature energy reservoir to a low-temperature energy reservoir according to the first law of thermodynamics. The maximum amount of heat converted to work is determined by the second law of thermodynamics. Whereas the clockwise Carnot Cycle produces work, a reversed Carnot Cycle (operating counterclockwise) acts as a heat pump, requiring work to transfer heat from a low-temperature energy reservoir. The basic principles of a reversed cycle are explained by Sadi Carnot in his paper *Reflections on the Motive Power of Fire*.

The reversing Carnot Cycle that uses heat supplied as an energy source and delivers mechanical work as the energy output, one will get reversed engine which applies mechanical work that is supplying as the energy source to transfer heat from a lower energy reservoir to a higher energy reservoir. The latter statement is the basis for refrigeration and AC.

Today, nearly all on- and off-highway vehicles manufactured in and for the United States, as well as a significant number of other world vehicles, have AC as standard equipment: passenger cars, trucks, buses, farm and construction equipment, and specialty vehicles, among others. Today, a vehicle without AC is similar to a passenger car without a heater during the early 1930s, when heaters were still just options.

The first mobile AC systems were introduced in the late 1930s by Packard, Chrysler, and Cadillac. The first supplier was Bishop & Babcock Company of Cleveland, Ohio. In the mid-1940s, AC units appeared in the aftermarket in the Southwestern United States and were installed by dealers in small shops. The first plant installing AC units opened in 1953. This was the first and only unit in which an evaporator in a passenger vehicle was located in the trunk and ducted from there to blow air forward. In 1954, the front units were brought to the market by Nash and Pontiac. Today, all passenger vehicles have an HVAC system located in the front of the vehicle, but most buses, truck sleeper cabs, vans, and other vehicles are equipped with a rear HVAC system.

Over the years, the HVAC system has been greatly improved in the design of heat transfer surfaces, plumbing system, compressor, and air distribution. Today, the HVAC system is not only of important value to the vehicle, but also, very often it is necessary equipment (e.g., for ambulances and fire trucks). Standards in a large variety of industries include specifications and requirements for an HVAC system as a necessary option.

In the first AC systems, all plumbing was done by copper tubing connected by commercial refrigeration fittings. The compressor was four cylinder and camshaft operated. Those systems' operational mode was 100% recirculation. A little later, fresh air/outside air capability was added to help improve the defrosting and demisting features significantly.

MOBILE HVAC BASICS: PHYSICS OF THE HVAC SYSTEM

The mobile HVAC system's main duty is to provide driver and passenger temperature comfort and the required cab dehumidification during hot, high-humidity weather without failure due to device cooling capability and while still allowing safe vehicle operation. It must cool down the vehicle enclosure within a short time, often within 30 min, at ambient temperatures up to 45°C–50°C. The dehumidification feature in great part determines the required system's (evaporator's) cooling capacity. The heating part of a mobile HVAC system should provide acceptable defrost, demist, and defog performance, yet the heater must make drivers and passengers comfortable in relatively short periods after cold startups at temperatures of −30°C or lower (in some cold areas of the United States or the world). The system must be quiet, and the controls must be easy to understand and operate.

HVAC systems have three fluid streams: air, coolant, and refrigerant. Each of these streams defines heat transfer between the corresponding heat exchanger and fluid. The automotive climate control system includes three main heat exchangers: the evaporator, condenser, and heater. An AC thermal system consists of the evaporator and condenser with two heat transfer fluid streams: refrigerant and air. A heating thermal system consists of the heater with two heat transfer streams as well: coolant and air. Each of these two thermal systems affects the human body's comfort level in the summer or wintertime, correspondingly. Heat transfer processes in the thermal system are defined as exchanges of heat between the heat receiver (heat sink), which is refrigerant in an AC thermal system, and air in a heat thermal system and the heat supplier (heat source), which is air in an AC system and coolant in a heat system.

The physics of the evaporating and heating processes can be described as follows.

Evaporating Process

An evaporator acts as a device to lower the temperature of ambient or recirculated air that passes through it. The physics of the process is explained in the following manner: the low-temperature and low-pressure liquid refrigerant enters the liquid line of an evaporator, where it is boiled (evaporated) by the air coming through (evaporation Line 1–3 of the reversed Carnot Cycle; Fig. 1). The evaporator must provide stable refrigerant flow under every possible operating condition and should have sufficient capacity for rapidly lowering the vehicle's cabin temperature.

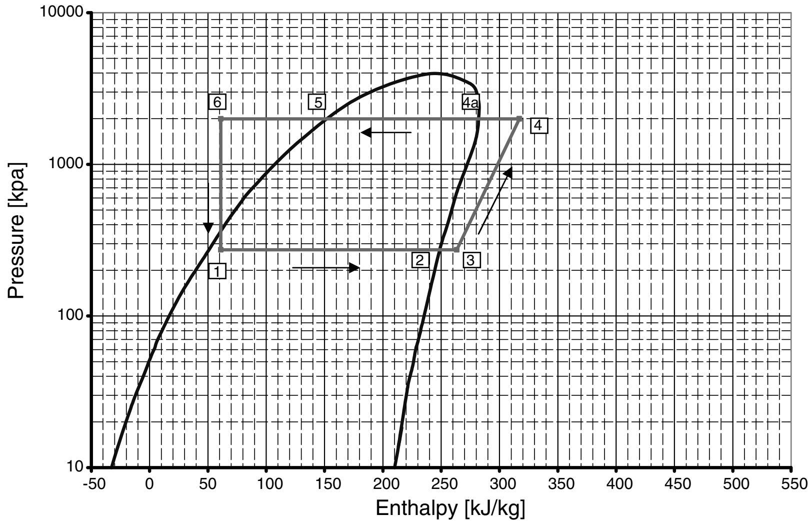


Fig. 1 Pressure–enthalpy diagram HFC-134a.

The features of the evaporation process are based on the theoretical principles of the reversed Carnot Cycle (Fig. 1), which actually represents the ideal refrigeration cycle. From the features of this cycle immediately follow the necessary conditions for the best-performance evaporation process, which are constant pressure and temperature of the refrigerant on the evaporation line of the pressure–enthalpy diagram (Fig. 1). In view of this statement, it is clear that all other things being equal, the same evaporator provides more capacity when the refrigeration-process conditions are closer to the ideal refrigeration process (or reversed Carnot Cycle) conditions.

The next very important consideration that greatly affects the evaporator's design is that the refrigerant's circuiting must be designed to make the outer surface temperature as uniform as possible.

It is imperative to explain here the features of the refrigeration cycle (reversed Carnot Cycle) for better understanding of all HVAC parts' assignments and duties. A compression-refrigeration cycle (Fig. 1) consists of the following phases: isentropic compression; condensation, including desuperheating, actual condensation, and subcooling; adiabatic expansion; and evaporation, including superheating (if any).

Compression (discharge Lines 3–4; Fig. 1) is accomplished by means of an AC compressor where superheated vapor enters the compressor at low pressure and temperature, and is compressed to a higher pressure and temperature. This task is performed at constant entropy in an isentropic process in which change in entropy is zero. The difference in enthalpy between the high-pressure and low-pressure gas is an amount of compressor work or a measure of available energy.

Condensation (Lines 4–6). The high-pressure superheated vapor is carried to the condenser, where the refrigerant undergoes the following three transformations (phases) utilizing outside ambient air: (1) the heat is removed from the gas refrigerant vapor to bring its temperature to a condensing point (desuperheating, Lines 4–4a); (2) gas is condensed to a liquid phase (actual condensation, Lines 4a–5), and (3) the liquid is subcooled to a specified level (subcooling, Lines 5–6). The difference between the enthalpy from desuperheating to subcooling is a measure of the heat content released to the ambient air.

The subcooled refrigerant liquid then undergoes expansion through an orifice device where it is expanded adiabatically (the change in enthalpy is zero; there is no heat exchange with surrounding), changing the phase to a low-temperature and low-pressure mixture of liquid and vapor gas (Lines 6–1).

The low-temperature and low-pressure mixture of liquid and vapor gas enters the evaporator, where the latent and sensible heat from the ambient or recirculated air passing through the evaporator core vaporizes (boils) the refrigerant into a gas phase (Lines 1–3). This process is performed at constant pressure and temperature. The difference in enthalpy between the liquid and gas refrigerant is a measure of the amount of heat removed from the air. This amount is defined by evaporator capacity. The evaporator also acts as a dehumidifier by condensing the water vapor (latent heat) from the air flowing through the coil. If the AC system is equipped with an expansion valve, either thermostatic (thermal expansion valve [TXV]) or electronic (electronic expansion valve [EXV]), the superheating line (Lines 2–3) continues the evaporation process. This system has to be equipped with receiver/dryer, located between the condenser and the expansion valve, where any moisture is removed from the refrigerant. If the AC system is an orifice tube (OT) system, the superheating line does not exist. The refrigerant exiting the evaporator has superheat equal to zero (a significant advantage of an OT system vs a TXV system). In an OT system, the entire evaporator surface is used to cool air, whereas in a TXV system, part of the evaporator surface is used to superheat gas (see the detailed comparison of TXV and OT systems in "System Configurations and Overview"). The OT system has to be supplied an accumulator located between the evaporator and compressor to prevent any liquid from entering the compressor.

The low-temperature and low-pressure refrigerant gas enters compressor through the suction line, and the cycle is repeated.

Heating Process

The heating process is less complicated than the evaporation process. The high-temperature liquid (usually, coolant from the engine's cooling system) enters the heater core, where it is cooled by the air coming through. The heater must provide stable coolant flow under every possible operating condition. The heater core must be large enough to ensure the thermal comfort of the cab crew while the windshield is defrosted and the cab area is defogged within the requirements and conditions defined by the Society of Automotive Engineers (SAE). The above-mentioned consideration of uniformity of the surface temperature, which greatly affects the evaporator design, influences the selection of the heater core, which also must be designed to make the outer surface temperature as even as possible.

One more point that should be mentioned pertains to packaging constraints. Although it seems that these constraints are not so important, for the automotive application, packaging constraints end up being the most decisive factors in the design of the heat exchangers. Because the evaporator and heater are part of HVAC module, they should be packaged as part of the whole unit. Space is very limited; therefore, the selection of the heat exchanger type in great part depends upon space availability.

The same note has to be made with regard to the frontal area of the device. The frontal area should be separated from the whole apparatus because on one hand, the surface area is a very great factor in influencing the evaporator and heater capacity—that is, the greater this area, the better the heat exchanger's performance. On the other hand, the frontal area of the heat exchanger is limited (again) by the space available for the device.

SYSTEM CONFIGURATIONS AND OVERVIEW

A mobile HVAC system consists of:

- H—heating system, which usually includes a heater core, a heater valve, controls, and plumbing
- V—ventilating part, which mostly contains an inlet cowl, ducts, a blower and blower housing, a blend air door (if any), mode doors, louvers, baffles, actuators, and a condenser fan with a fan shroud and seals
- AC—AC system that incorporates the condenser, evaporator, compressor, controls, expansion device, accumulator or receiver/dryer, and plumbing

The environment in which the vehicle operates sets the HVAC requirements. The heat load of the environment on the vehicle HVAC system consists of heat conduction into the crew/passenger compartment, air infiltration, heat pickup of the incoming air to the evaporator, sun load (more than 50% of total), driver's and passengers' body heat, and the heat (sensible and latent) of the incoming air. All the above-mentioned factors affect AC system performance. Similar considerations can be brought up with regard to the heating system's performance, of course, bearing in mind the low ambient temperature that is required to warm the enclosure. Human-body and sun-load heat in an HVAC system's heating mode help warm up the enclosure.

In Figs. 2 and 3, the TXV and OT mobile HVAC system schematics are presented. To familiarize readers with HVAC components and their operations, the following lists all these components.

- Compressor
- Condenser
- Receiver/dryer
- Expansion valve
- Orifice tube
- Accumulator
- Evaporator
- The supporting cast

How does an AC system work? As we have already discussed, we cannot create cold. The only means of reducing temperature is removing heat and maintaining enclosure temperature and humidity level. An AC system contain two sides: a high side (high pressure and

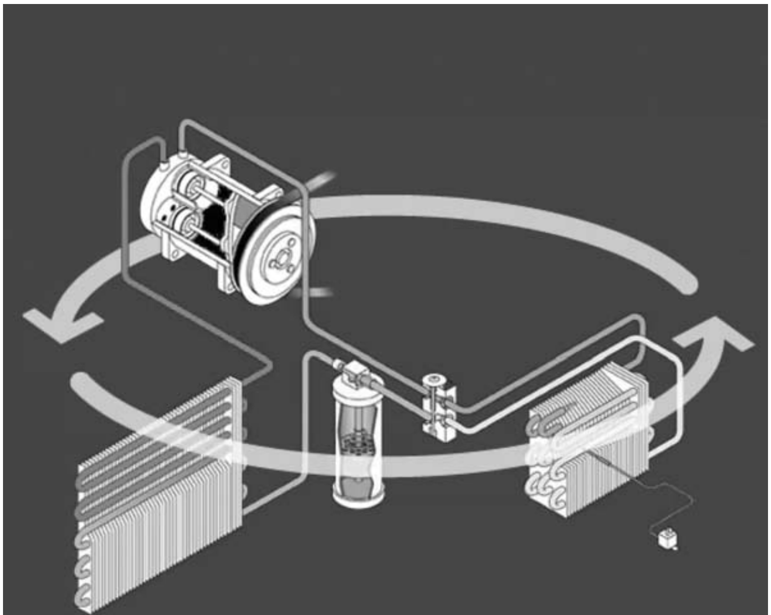


Fig. 2 TXV system.

Liq–Mobil

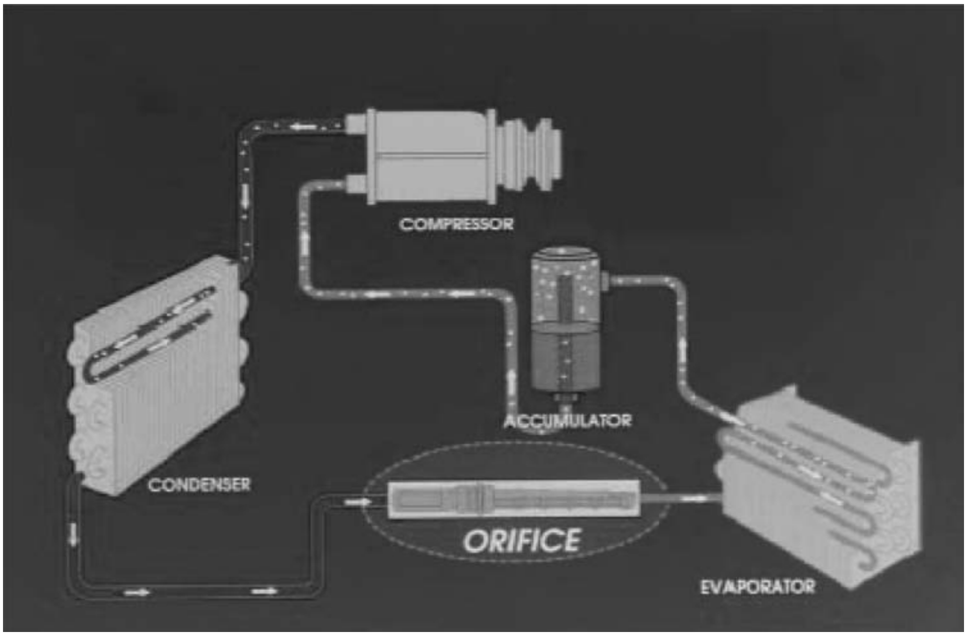


Fig. 3 OT system.

temperature) and a low side (low pressure and temperature). High side: from compressor discharge outlet to condenser to receiver/dryer (TXV system) to expansion device inlet. Low side: expansion device outlet to evaporator coil to accumulator (OT system) to compressor inlet.

SYSTEM COMPONENTS AND DUTIES

Compressor

The primary types of compressors used in the mobile HVAC industry are piston (most usable; see Fig. 4), rotary (work like rotary engines), and scroll (work like superchargers).

The compressor is the heart of the AC system. Its primary function is to compress cool, low-pressure vapor and pump out high-pressure hot vapor. The compressor draws low-pressure, low-temperature refrigerant gas from the evaporator. Most mobile AC compressors are engine-driven devices. The higher the engine's RPM, the greater the compressor output. New electrified mobile HVAC systems equipped with electrical hermetic compressors are described in the entry "Mobile HVAC Systems: Fundamentals, Design and Innovations", of this Encyclopedia.

The main cause of compressor failure is lack of lubrication. A normal rule about the required amount of oil in refrigerant is about 200–250 g of oil for every 1 kg of refrigerant in the system. The refrigerant oil level must be maintained for proper lubrication and sealing of the compressor.

Fig. 4 Piston type compressor.

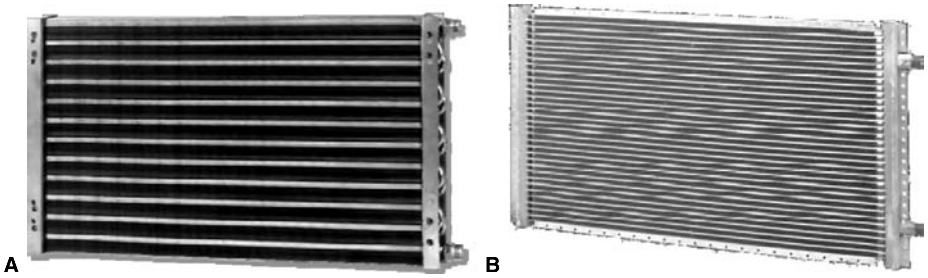


Fig. 5 (A) Tube and fin condenser and (B) parallel flow condenser.

Condenser

The two major types of condensers are tube and fin (Fig. 5A) and parallel flow condensers (Fig. 5B). There are also serpentine-type condensers, which usually are not used as often as the other two types. Primary condenser functions are to release enough heat to the outside air to cool the refrigerant from a hot gas to a less-hot liquid. The condenser must have air blowing across it at all times by means of a fan and fan shroud.

The condenser allows for the removal of heat or energy from the truck cab or the vehicle's inside compartment to provide a more comfortable less humid environment.

As mentioned above, while flowing through the condenser, refrigerant changes its phase from vapor to liquid via the removal of heat into the cooler outside air. Refrigerant R134a, which is used mostly in automotive applications, changes its vapor phase to liquid phase at 49°C at a constant pressure of 1172 kPa. The condenser actually operates in three modes: desuperheating the inlet refrigerant gas, condensing the saturated vapor, and subcooling the liquid refrigerant. The condensing portion requires the greatest surface area, followed by the other two areas.

Under normal operating conditions, the temperature differential between the refrigerant going into the condenser and coming out of the condenser is about 5°C. The condenser is usually trouble free, but two problems may occur. Cracks and dents, which are nonrepairable, cause leakage and external blockage of the fins via debris that reduces the capability of the fins to cool properly.

Receiver/Dryer

The TXV system (Fig. 2) is equipped with a receiver (filter)/dryer. The primary functions of the receiver (filter)/dryer (Fig. 6A) are to filter liquid refrigerant, to act as a moisture remover (to absorb moisture), and to serve as a temporary storage unit for liquid refrigerant until it is needed by the system. There are two types of dryers: bag (Fig. 6B) and water (the most common; Fig. 6C). The basic parts of the receiver/dryer are shown in Fig. 6C. The necessity of removing moisture from refrigerant stems from the fact that a chemical reaction between refrigerant and water creates acid—in particular, R134a + water = hydrofluoric acid. Under normal operating conditions, the temperature differential between the refrigerant going into the dryer and coming out of it is about 1°C–2°C. If frost occurs on the dryer when the system is in operation, it becomes plugged, and it will need to be replaced.

Expansion Valve

A TXV device meters refrigerant flow to the evaporator and protects the compressor from liquid refrigerant. The

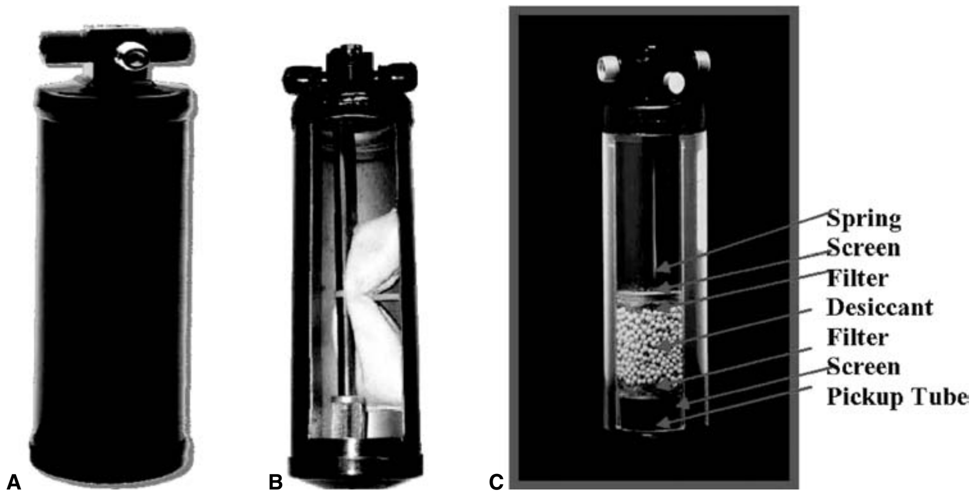


Fig. 6 (A) Receiver/Dryer, (B) bag type receiver/dryer, and (C) water type receiver/dryer.

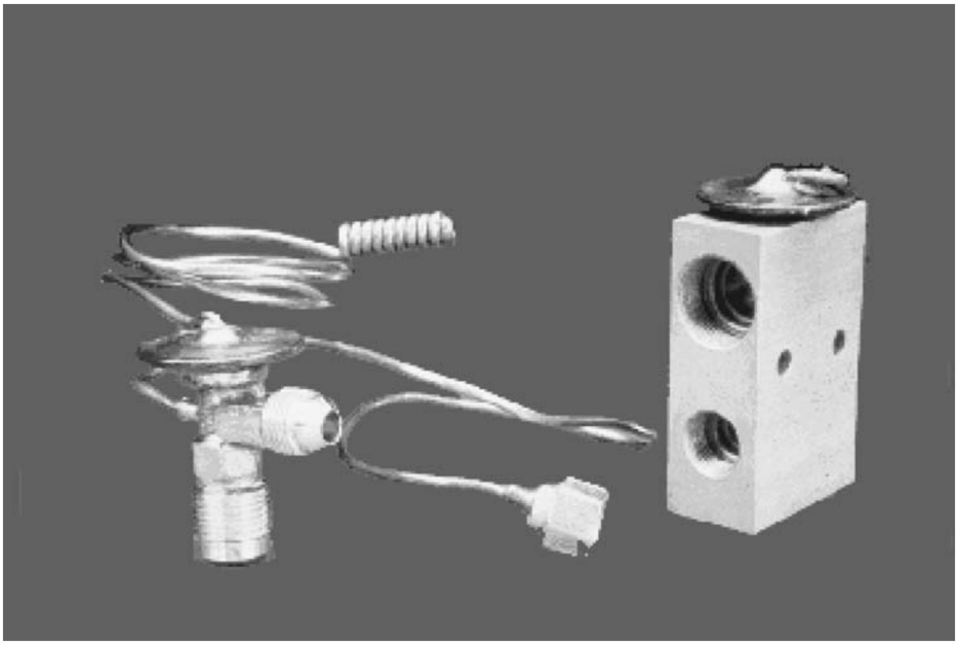


Fig. 7 (A) Bulb type or externally equalized TXV and (B) block type or internally equalized TXV.

schematic of a TXV system is shown in Fig. 2. There are two types of TXV devices: capillary bulb (externally equalized; Fig. 7A) and block (internally equalized; Fig. 7B).

The expansion valve controls the amount of refrigerant sprayed into the evaporator; otherwise, liquid refrigerant would be allowed to enter the compressor, which could cause severe damage. The valve responds to changes in heat load conditions, and actually, TXV operates in response to superheat. Increased heat load causes the valve to open, which allows more refrigerant to flow through. A thermostatic bulb senses the evaporator outlet tube's temperature (externally equalized) or the evaporator inlet tube's temperature (internally equalized). The medium in the bulb (usually, refrigerant) exerts pressure on the diaphragm. An increase in temperature (heat load) causes the diaphragm to open, thereby increasing flow. A decrease in temperature causes the diaphragm to close, decreasing refrigerant flow accordingly. Under normal operating conditions, the valve is hot at the inlet and cold at the outlet, and the temperature differential between refrigerant going into the valve and coming out of it is about 4°C–6°C.

A TXV device holds a preset superheat by modulating the refrigerant flow mechanically via a diaphragm, whereas an EXV modulates refrigerant flow electronically, using pulse width modulation or a stepper motor. The superheat can be controlled between 0.5 and 11°C. This device can maximize the AC performance; however, the EXV is significantly more expensive than TXV and less reliable, which can cause an increase in warranty costs.

Orifice Tube

A schematic of an OT system is presented in Fig. 3. The orifice tube (Fig. 8) is nothing more than a piece of copper tubing. This device is a calibrated tube that meters the liquid refrigerant and regulates its flow to satisfy evaporator demand. OTs are color coated to ensure proper replacement. The orifice flow rate is affected by the amount of subcooling out of the condenser. The greater the subcooling, the higher the refrigerant flow rate. Because the OT size is fixed, its size is optimized by maximizing the system performance at high, moderate, and low evaporator load at appropriate compressor speed. Under normal operating conditions, the tube is hot at the inlet and cold at the outlet. There is a large temperature differential between refrigerant going into the OT and coming out of it—about 30°C–80°C. The inlet of the OT is equipped with a screen to protect the evaporator and the compressor from debris and from contaminated refrigerant. Blockage of this screen could cause failure of the whole AC system, because the system will be cycling very rapidly.

A variable orifice tube (VOV) differs from fixed OT in that it is capable of varying the refrigerant flow restriction through the tube in response to the refrigerant temperature exiting the condenser. (During vehicle idling time, the refrigerant temperature is high, and the restriction is greatest; at vehicle highway speed, the refrigerant temperature is low, and the restriction is lowest.) VOV expansion is not used much because its advantages over fixed OT are not significant, and the cost is much higher.

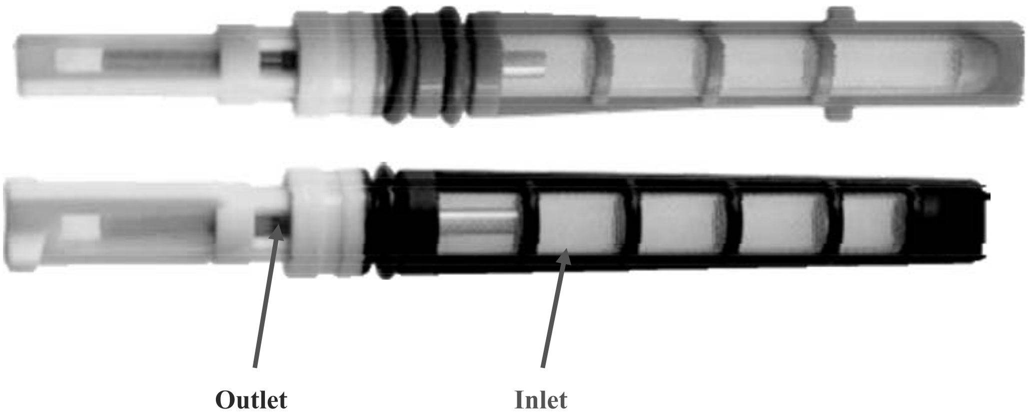


Fig. 8 Orifice tube.

The advantages and disadvantages of an OT system vs a TXV system will be discussed later in this section.

Accumulator

An OT system (Fig. 3) is equipped with an accumulator. Accumulators (Fig. 9) vary in style, but they all do the same job and do it the same way. The accumulator is a liquid vapor separator that stores (accumulates) excess refrigerant that comes out of evaporator so that it does not reach the compressor. Usually, accumulator contains desiccant that dries refrigerant by absorbing moisture from it. The accumulator also filters refrigerant. When the accumulator operates normally, it will be cold.

Evaporator

The evaporator is a refrigerant/air heat exchanger in which the liquid refrigerant absorbs the heat that is transferred from the air passing over the evaporator. A blower pulls cab or fresh air through the evaporator (Fig. 10B), where the refrigerant absorbs heat in the air. The cool air is pushed back into the cab and either blows free or distributed through the ductwork. So the evaporator's duties are to absorb the heat, dehumidify the air, and filter the air. A large number of evaporator types exists in the mobile AC industry. The four basic types of evaporators are: plate fin; tube and fin (Fig. 10A); serpentine; header tube and spacer. There are other types of evaporators as well. In the entry "Mobile HVAC Systems: Fundamentals, Design and Innovations", of this Encyclopedia the evaporator capacity calculation and design-optimization approaches will be discussed.

Under normal operating conditions, evaporator in and out lines are cold, and the temperature differential between refrigerant going into the coil and coming out of it is about

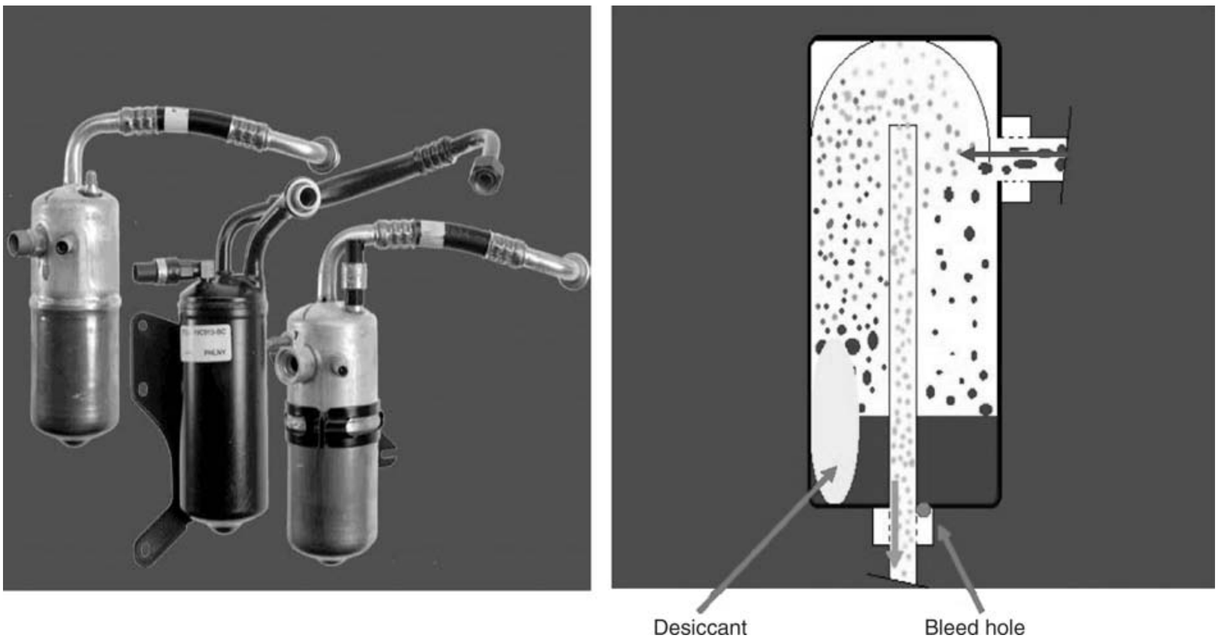


Fig. 9 Accumulators.

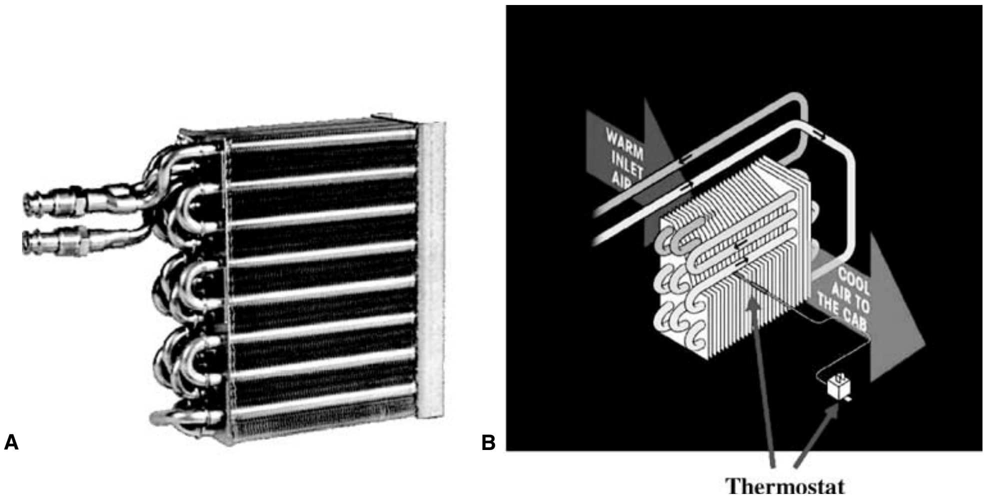


Fig. 10 (A) Tube and fin evaporator and (B) evaporator performance schematic.

3°C–6°C. Basically, the evaporator confirms that the AC system is performing to expected specifications by measuring the air temperature as it leaves the heat exchanger.

The Supporting Cast

Thermostats

Thermostats (Fig. 11A) sense the evaporator surface and protect coils from freezing (also see Fig. 10A). The three basic types of thermostats are rotary, cable controlled, and fixed setting. Thermostats turn the clutch on and off by sensing evaporator temperature. A probe inserted into the evaporator (Fig. 11B) measures evaporator surface temperature (only the last 0.1 m of the tube is sensed) and turns the clutch on and off (disengaging and engaging the compressor) by sensing this evaporator temperature. The tip must be close to the coldest spot of the evaporator. This spot is usually determined by testing the unit on the test bench (calorimeter). The location of the coldest spot depends on the airflow rate and air profile, HVAC unit configurations, and evaporator design.

Pressure Switches

Most of the AC systems have High and Low pressure switches (Fig. 12) that protect compressor/system against too low or too high pressure by disengages compressor at set low or set high pressure accordingly. Some vehicles have a third high-side pressure switch for cycling condenser fan to cool the condenser.

Electrical System Basics

An HVAC system electrical circuit includes On/Off switch, Thermostatic switch, Pressure switches (Low-side low pressure, High-side low pressure, High-side high pressure), Compressor Clutch. Recently a few electronic compressor/system protection devices have been introduced (Fig. 13).

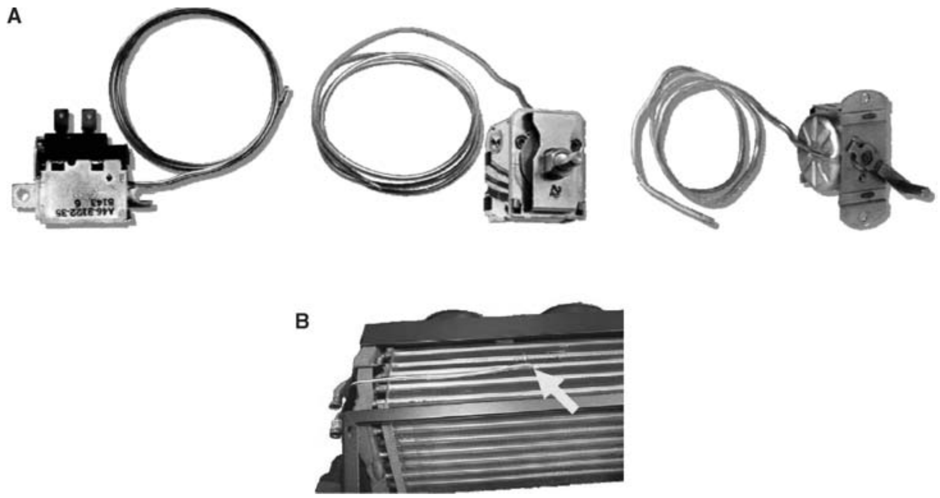


Fig. 11 (A) Rotary, cable controlled, and fixed setting thermostats and (B) evaporator with installed thermostat's probe.

Liq–Mobil

Fig. 12 Pressure switches.

Fig. 13 Compressor / AC system protection devices.

These devices protect the AC system against low and high pressures, sense low refrigerant in the system, identify low voltage, measure some other refrigerant system parameters, and generate special codes when any of those parameters goes to a dangerous range and shut off system accordingly.

A few other supporting parts are fittings, hoses, and water/heater valves.

TXV System vs OT System

The expansion valve is a much more complex device than the orifice tube, and it offers a wider range of control. However, our experience shows that this advantage of the TXV system is overridden by many problems that arise. These problems are listed below. The expansion valve is the leading source of AC system failure and, thus, the highest warranty item. An OT system allows the use of all evaporator surfaces for heat transfer, whereas a TXV system uses a significant part of the evaporator surface for superheating refrigerant. The expansion valve controls the flow of the liquid refrigerant entering the evaporator relative to the amount of superheat in the refrigerant gas leaving the evaporator. Because the TXV operates in response to superheat, part of the evaporator must be used to superheat the refrigerant gas. The higher the superheat (0.5°C–11°C), the more of the evaporator surface is used for superheating.

Another problem that arises in a TXV system is so-called system hunting. The system goes rapidly on and off, which alternately starves and overfeeds the evaporator. There can be multiple reasons for hunting: A valve is oversize; the factory-set TXV superheat is too low; the suction line is too small; and so on. Our experience and apple-to-apple testing results show that the OT/accumulator-design systems allow up to 2°C–3°C better performance on the air side over TXV design. (The OT system has a temperature differencial between evaporator air in/out that is 2°C–3°C better.) Last but not least, the advantage of the OT system is that it takes more charge because of accumulator use, which allows a longer period between system required to add charge due to permeation through hoses and other small leaks.

Helpful Tips

A few helpful tips related to mobile HVAC systems:

- To get mobile HVAC system certification, contact the Mobile Air-Conditioning Society (MACS), P.O. Box 88, Lansdale, PA 19446, (215) 631–7020, www.macs.org.
- To get information related to mobile refrigerants or other useful topics, contact Refrigerants EPA, (800) 296–1996, www.epa.gov/ozone.

ACKNOWLEDGMENTS

The author wishes to acknowledge the support of ACC Climate Control, Inc., which allowed him to use materials and documents that are the property of ACC Climate Control, Inc. Also, the author would like to express profound gratitude to Al Kargilis, P.E., whose manuscript[2] was used extensively during the preparation of this entry.

REFERENCES

1. Hewitt, G.F.; Shires, G.L.; Polezhaev, Y.V. *International Encyclopedia of Heat & Mass Transfer*; CRC Press: Boca Raton, FL, 1996.
2. Kargilis, A. *Design and Development of Automotive Air Conditioning System*; University Consortium for Continuing Education: Encino, CA, 1997.

National Energy Act of 1978

Robert R. Nordhaus
Van Ness Feldman, PC, Washington, D.C., U.S.A.

Abstract
The National Energy Act (NEA) of 1978 was passed by U.S. Congress in response to the energy crisis of the 1970s. It was designed to resolve a disjointed national energy policy and empower the United States with greater control of its national energy destiny. The NEA and its progeny established energy efficiency programs, tax incentives, tax disincentives, energy conservation programs, alternative fuel programs, and regulatory and market-based initiatives. Results of the NEA have been mixed. Most of the market-based initiatives have been retained, but many of the regulatory initiatives have since been abandoned.

On April 20, 1977, President Jimmy Carter submitted his National Energy Plan (NEP) to Congress—113 specific legislative and administrative proposals. Congress responded by ultimately passing five pieces of legislation, collectively known as the National Energy Act of 1978 (NEA): the Natural Gas Policy Act (NGPA)[1]; the Public Utility Regulatory Policies Act (PURPA)[2]; the Energy Tax Act[3]; the Powerplant and Industrial Fuel Use Act (Fuel Use Act)[4]; and the National Energy Conservation Policy Act (NECPA).[5]

THE PRESIDENT'S PROPOSALS IN CONTEXT

Energy legislation in the United States since 1970 has been closely tied to oil. By the beginning of the decade, the United States had become a net importer of oil. In 1973, the Organization of Petroleum Exporting Countries (OPEC) imposed an oil embargo against the United States, reducing the availability of foreign crude and throwing world petroleum markets into turmoil. By 1985, its delivered price to U.S. ports had increased fivefold. Congress and every administration throughout the decade responded by passing legislation. Thus, by 1980, our national energy policy comprised over 100 different, interrelated (and sometimes inconsistent) federal policies and programs.

By 1977, oil imports stood at 50% of U.S. oil consumption, and the general consensus was that the U.S. energy situation had grown to "crisis" proportions. As one House Committee stated: "The fundamental problem for U.S. Energy Policy is the insecurity of its oil supply… There is widespread agreement that a continuation of present trends will lead to a very tight world oil market in the mid-1980s…The United States faces the problem of making the transition from an era of cheap abundant energy to relative scarcity of expensive energy supplies."[6]

It was in this context that President Carter submitted his NEP, which proposed both demand-side and supply-side initiatives. Carter's proposal embraced three broad objectives for the country to regain control of its energy destiny: (1) in the short term, reduce dependence on foreign oil and limit supply disruptions; (2) in the medium term, prepare for the eventual decline in the availability of world oil supplies caused by capacity limitations; and (3) in the long term, develop new, reliable, and inexhaustible sources of energy for sustained economic growth.

To achieve these objectives, the NEP contained three major components: first, an ambitious conservation program for all sectors of energy use to reduce the annual rate of growth of demand; second, measures to induce industries and utilities using oil and natural gas to convert to coal and other abundant fuels; and third, a national pursuit of a vigorous research and development program to provide renewable and other resources to meet domestic energy needs. Major features of the NEP included conservation and fuel efficiency; rational pricing and production policies; reasonable certainty and stability in government policies; substitution of abundant energy resources for those in short supply; and development of nonconventional technologies for the future.

For oil and natural gas, the NEP intended to provide prices that encouraged production from new fields and established a more rational pattern of distribution, while preventing windfall profits. It also intended to promote conservation by confronting oil and gas users with increased prices. The NEP called for expanding the Strategic Petroleum Reserve to one billion barrels, diversifying sources of oil imports, and accelerating the development of contingency plans to reduce U.S. vulnerability to foreign oil supply interruption.

Keywords: National Energy Act; National Energy Plan; Utility; Regulation; Legislation; Natural gas; Fuel use; Energy conservation; Energy tax.

In the transportation sector, the NEP proposed initiatives to reduce the demand for gasoline and other transportation fuels. In energy efficiency, the NEP proposed a reduction of energy wasting in existing buildings, an acceleration in the development of mandatory energy efficiency standards for new buildings, and the establishment of mandatory minimum energy efficiency standards for major appliances. In the electric industry, the NEP proposed the removal of major institutional barriers to cogeneration and to provide an additional tax credit for investment in cogeneration equipment. To promote industrial conservation and fuel efficiency, the NEP proposed an additional tax credit for energy-saving investments (including solar energy equipment) and conservation retrofits of buildings. Lastly, the NEP contained a broad program for utility reform.

The NEP's reception on Capitol Hill was rocky, but after 18 months of debate, the legislation—now in five pieces—was enacted. The components are described below.

THE NATURAL GAS POLICY ACT OF 1978

In 1954, the Supreme Court held that producer sales of natural gas into the interstate market were subject to the rate and certification requirements of the Natural Gas Act (NGA). When prices in the intrastate gas markets raised above the rates the Federal Power Commission (FPC) set for interstate producer sales in the 1970s, gas producers directed most new gas into the intrastate market, creating recurrent supply shortfalls in the interstate market. These supply shortfalls were exacerbated by the extraordinarily cold winter of 1976–1977 when serious natural gas shortages occurred in regions served by interstate gas pipelines. That winter helped shape the debate in 1977 and 1978 on President Carter's natural gas proposals. In contrast to the oil embargo whose cause was economic and political, the natural gas crisis was seen, at least in part, as a regulatory problem. Critics of the FPC blamed the crisis on the low price it set for interstate gas. Others argued that the United States was wasting clean, valuable gas on low-priority uses, such as boiler fuel. As the NEP explained, "As a result of regulation under the [NGA], natural gas is now substantially underpriced, and there is excess demand. Existing supplies are being wasted on nonessential industrial and utility uses…[The] intrastate-interstate distinction has also become unworkable, indeed intolerable, as the limited amount of new gas increasingly flows to the unregulated intrastate market at the expense of interstate consumers."[7]

The Carter administration bill went through nearly 18 months of debate in Congress, pitting members from producing and consuming states against each other. Members from consuming states wanted price controls extended to all gas (both interstate and intrastate), while members from producing states wanted deregulation of interstate gas and opposed imposing price controls on intrastate gas.

The NGPA proved to be the most controversial portion of Carter's energy program. The final product was an intricate compromise that retained most of the elements of the President's proposals, in particular the extension of controls to the intrastate market, so that all gas produced in the United States was subject to price controls. It also included a key element of the Senate proposal—phased deregulation of the wellhead price of new natural gas. New natural gas and most other gases were deregulated between 1979 and 1987. The NGPA also included an "incremental pricing" mechanism that was designed to allocate the costs of expensive new gas to industrial (rather than residential) users.

WELLHEAD PRICING

Natural Gas Policy Act's wellhead pricing system was designed to encourage the exploration and production of new natural gas through a series of complex pricing formulas. In effect, gas from new wells or new fields was priced at levels higher than gas from existing wells. The Federal Energy Regulatory Commission (FERC), successor to the FPC, was empowered to regulate all "first sales" of natural gas. A "first sale" was essentially any sale or transfer for value, except sales by pipelines or distributors (other than from their own production). The price that could be obtained by the producer for a first sale varied according to the classification of the gas in the wellhead price determination procedure. In most cases, state, not federal, agencies were to categorize producing wells to determine the appropriate pricing category.

Prior to the NGPA, producing states (other than New Mexico) did not impose price restraints on producer sales of natural gas, and federal price controls on producers extended only to interstate wholesale sales. As noted above, under pre-NGPA regulation, the FPC's prices for gas sold in the interstate market were lower than unregulated prices in the intrastate market. As a result, interstate pipelines were unable to acquire enough supply for resale. The NGPA remedied this, imposing the same ceiling prices on gas newly committed to the interstate and intrastate markets (Gas already committed to one market or the other generally remained subject to a ceiling price based on pre-NGPA prices.). Federal wellhead pricing was terminated in 1989.

INCREMENTAL PRICING

The incremental pricing provisions of the NGPA were designed to shelter high-priority users from the higher gas prices allowed by the NGPA, and to price gas to low-priority users at the alternative fuel price. However, these provisions never had any practical effect because of changing market conditions, and in May 1987, Congress repealed them.

CURTAILMENT

In the NGPA, Congress played a direct role for the first time in establishing a system of curtailment priorities to allocate the gas supplies of interstate pipelines in their supply shortages. Because of the improved supply situations of the interstate pipelines, these provisions also had little practical effect.

EMERGENCY AUTHORITY AND TRANSPORTATION

The grant of presidential authority in energy emergencies, authorized in Title III of the NGPA, was a substantial reenactment of the Emergency NGA, enacted early in 1977. However, the transportation authority Title III conferred upon FERC represented an important departure from prior law. Specifically, section 311 of the NGPA[8] gave FERC the authority to authorize, by rule or order, any interstate pipeline to transport gas on behalf of any intrastate pipeline or local distribution company (LDC) and to authorize any intrastate pipeline to transport gas on behalf of any interstate pipeline or LDC, all without the need for a certificate, under section seven of the NGA. Transportation so authorized would not render any person subject to the jurisdiction of the NGA. This section, according to the House's Committee, gave FERC the power to "facilitate development of a national natural gas transportation network without subjecting intrastate pipelines, already regulated by State agencies, to [FERC] regulation over the entirety of their operations."[9]

Section 311 can be viewed as an important step toward the restructuring of the gas pipeline industry in the 1980s, which permitted pipelines to render unbundled transportation services (i.e., transportation of gas separate from the sale of the commodity). Although unbundled transportation services had been authorized under prior law (e.g., FPC Order No. 528),[10] FERC's section 311 rule authorized transportation services on a "self-implementing basis," that is, no prior regulatory approval of the particular transportation service was required.[11] This became a model for FERC's later blanket transportation rules and Order No. 436 program under the NGA.

Today, of course, the importance of section 311 transportation has been overshadowed by FERC Order No. 636,[12] in which FERC largely took interstate pipelines out of the business of selling gas, requiring them to unbundle and offer transportation service to all customers on a nondiscriminatory basis. FERC has retained utility-type regulation over the interstate transportation function, while allowing the market for gas commodity itself to develop competitively.

It was in this context that FERC began a decade of restructuring the gas industry, issuing a series of orders that brought competition to the pipeline industry and allowed vastly increased customer choice. The NPGA, and FERC policies that implemented the NGPA, were the source of this restructuring. Phased deregulation, integration of the interstate and intrastate markets, and self-implementing transportation were the keys to creating the competitive gas supply market of the 1990s.

THE PUBLIC UTILITY REGULATORY POLICIES ACT

The PURPA is discussed under that heading at p. 1201

ENERGY TAX ACT OF 1978

The President's proposal to Congress relied heavily on tax incentives and disincentives to elicit new energy production and discourage inefficient uses of price-controlled oil and gas. Not unpredictably, the tax incentives fared better than the disincentives in the legislative process. Congress adopted most of the incentives. However, with minor exceptions, none of the tax disincentives survived the legislative process.

TAX INCENTIVES

The NEA, as introduced, provided for residential energy credits (to encourage individuals to make energy conservation investments) and business energy credits (to encourage business investment in more efficient commercial and industrial facilities and equipment). It also contained tax incentives for geothermal production and changes in alternative minimum tax treatment of oil and gas intangible drilling expenses.

As enacted, the Energy Tax Act of 1978 provided residential tax credits to encourage conservation and business tax credits to encourage energy conservation investments and the production of renewable energy, alternative fuels, and certain high-cost natural gas.

TAX DISINCENTIVES

As introduced, the NEA also contained a series of taxes designed to impose a standby gasoline tax to increase end-user prices of price-regulated crude oil to market levels and to impose additional taxes on industrial users of oil and gas that were designed to shift consumption to other fuels.

First, imposition of a standby gasoline tax at up to 50 cents per gallon was proposed if national targets for reduction in gasoline consumption were not achieved. Second, the President proposed a Crude Oil Equalization Tax (COET) that would have imposed a tax equal to the difference between the ceiling price for price-controlled crude oil and the average cost of crude to refineries. Third, consumption taxes on industrial and utility uses of petroleum products and natural gas were proposed. The petroleum tax, which would be in addition to COET, was to be phased in and capped at $0.50 per MMBTU. The natural gas tax after phase-in was designed to offset the price difference between price-controlled natural gas and distillate fuel oil. Finally, a gas-guzzler tax was proposed on low-fuel-efficiency vehicles.

Except for the gas-guzzler tax, Congress did not accept any of these tax disincentives. The standby gasoline tax was wildly unpopular with the public and did not survive the House. The industrial oil and gas consumption taxes were not enacted. Crude oil equalization tax fell by the wayside in the Senate. It was uniquely unpopular with the oil industry because the administration had proposed both to retain price controls on the industry and increase its taxes. The experience with COET was instructive, however. Two years later, the administration reformulated the tax as a Windfall Profits Tax designed, among other things, to tax gains accruing to producers after decontrol of oil prices. That tax was enacted as the Crude Oil Windfall Profits Tax Act of 1980.

THE POWERPLANT AND INDUSTRIAL FUEL USE ACT

The Fuel Use Act was designed to reduce dependence on oil and petroleum products by providing for expanded use of alternative energy sources by the nation's electric power plants and major industrial installations. The Fuel Use Act prohibited these facilities from using petroleum or natural gas as a primary energy source unless an exemption was obtained from the Department of Energy (DOE). Power plants and major fuel-burning installations had to comply with detailed criteria to secure an exemption from the Act's prohibitions.

The self-described purpose of the Fuel Use Act was to conserve natural gas and petroleum for uses other than electric utility, industrial, or commercial generation of steam or electricity for which there are no feasible alternative fuels or raw material substitutes. It was designed to prohibit or, as appropriate, minimize the use of natural gas and petroleum as a primary energy source and "to conserve such gas and petroleum for the benefit of present and future generations." In other words, when the nation is seeking to solve the problem of a shortage of domestic oil and natural gas, the answer is to require large consumers to substitute other fuels.

In many respects, the Fuel Use Act superseded the largely-ineffective Energy Supply and Environmental Coordination Act of 1974 (ESECA).[13] Energy Supply and Environmental Coordination Act was an emergency program designed by Congress to foster greater coal utilization by power plants and major fuel-burning installations.

To the Carter administration, the NEA provided an opportunity to provide a more effective fuel-switching program than that provided by ESECA. The Fuel Use Act was intended to provide for a swifter conversion process. To its critics, it was a continuation of misguided policy that mistakenly predicted continuing shortages of oil and gas and perpetually higher prices of those commodities.

At the time of its passage, the Fuel Use Act was described as a bold experiment. As might be expected, however, when the apparent shortages in gas and petroleum products eased, enormous opposition to the Act grew. The Fuel Use Act discouraged industrial and large-volume gas use that artificially kept the demand for gas low. The Act prevented the gas industry from expanding its industrial customer base, which in turn prevented the allocation of fixed system costs to a larger number of customers; denied pipelines the opportunity to reduce take-or-pay obligations for undertaken gas; and prevented the construction of new facilities that would use gas as the primary fuel source, thereby denying plant operators and electric utilities the opportunity to take advantage of the short lead time required to construct a gas-fired facility. Moreover, although the Fuel Use Act contained an exemption provision, it had taken DOE from five to twelve months to act on requests for exemption. Therefore, when the statute was repealed in 1987, the repeal encountered very little opposition.

THE NATIONAL ENERGY CONSERVATION POLICY ACT

The theme of conservation is pervasive in the NEP. As President Carter stated to Congress when he submitted his energy proposals in 1977: "[T]he cornerstone of our policy is to reduce demand through conservation. Our emphasis on conservation is a clear difference between this plan and others which merely encouraged crash production efforts. Conservation is the quickest, cheapest, most practical source of energy."[14]

The prior administration and Congress had attempted energy conservation measures, as well. The principal statute, the Energy Policy and Conservation Act of 1975 (EPCA), had provided fuel-efficiency standards for automobiles and authorized the FEA to prescribe energy-efficiency standards for other consumer products. The NEP included specific conservation incentives targeted to residential consumers and businesses. In the area of buildings, the President proposed a national residential energy conservation program for existing buildings, highlighted by a tax credit for amounts spent on residential energy conservation investments. Utilities were to be directed to offer energy conservation services financed by loans that would be repaid through utility bills. For the low-income sector, direct grants totaling almost $530 million would be provided for conservation investments.

There were other loans and grant programs, including mandatory efficiency standards for new buildings, a federal building conservation plan, and a program to demonstrate solar energy on select federal buildings that would be retrofitted to reduce energy consumption.

The NEP also included a change in the investment tax credit to encourage businesses to invest in energy conservation and renewable energy technologies. Indeed, the proposals for cogeneration, district heating, utility rate reform, and taxes were all seen as part of the overall conservation program.

As passed, the NECPA promoted three major roles for the government in energy conservation: setting energy-efficiency standards designed to cut energy consumption in particularly energy-intensive products or uses; disseminating information about energy conservation opportunities; and improving the efficiency of federal buildings (thereby cutting the government's own energy bill). The surviving portions of NECPA included a residential program for low-income weatherization assistance, grants and loan guarantees for energy conservation in schools and hospitals, and energy-efficiency standards for new construction. Initiatives in these areas continue today.

In other areas, the implementation of NECPA proved more problematic. For example, NECPA strengthened the appliance energy-efficiency standards program established under EPCA by requiring energy-efficiency standards for thirteen types of appliances, assuming such standards were economically justified. But in 1982, DOE concluded that the energy standards envisioned under NECPA could not be economically justified. Years of litigation and subsequent action by Congress were required before appliance energy-efficiency standards would be established. In 1987, Congress adopted the National Appliance Energy Conservation Act (NAECA),[15] specifying energy-efficiency standards and proposals for periodic updates of such standards for a variety of major household and commercial appliances. National Appliance Energy Conservation Act was amended in 1988[16] and again in 1992[17] through the EP Act. Today, efficiency standards have been prescribed for all major categories of consumer products. The standards program is currently regarded as one of the key policy tools for reducing greenhouse gas emissions.

CONCLUSION

President Carter wanted the United States to regain control of its energy destiny and his administration developed an extensive plan for doing so. While Congress did not embrace all of President Carter's proposals, many of Carter's proposals did become law. Whether the legislation accomplished the goals of its drafters is another question. Except in the area of conservation, many of the regulatory initiatives that flowed from the NEA have since proven unworkable or no longer necessary and have been abandoned. For example, the extension of price ceilings to the intrastate natural gas market through the NGPA was an effective mechanism for relieving supply shortages at the time, but is no longer relevant because Congress found that price controls in both the interstate and intrastate markets were unnecessary and phased them out.

On the other hand, most of the market-based initiatives that flowed from the NEA have been retained. The NGPA laid the foundation for the deregulation of natural gas and that task has now been completed with the 1989 enactment of the Natural Gas Wellhead Decontrol Act. The removal of the interstate/intrastate market dichotomy under the NGPA, together with the new gas transportation authority contained in the statute, created a unified national market for gas, eased the supply acquisition problem, and helped remove market distortions. The result was a more efficient allocation of natural gas resources and, after the phased deregulation of wellhead prices, it is probably the greatest contribution of the NGPA. Moreover, the most lasting effect of the NGPA, wellhead decontrol, led to a critical economic change: natural gas has now become a separate and distinct economic commodity. By removing both price and nonprice regulation over most producer sales of natural gas, the NGPA cleared the way for natural gas to be sold as a commodity, unbundled from transportation, by a wide variety of market participants, without regulatory approval.

The lessons learned from the NEA's tax initiatives remain applicable today. Tax policy in the NEA had an asymmetric quality; Congress would provide credits or deductions, but would not impose consumption taxes. This is still the case, as witnessed from the outcry in 1993 when President Clinton proposed a Btu tax.

Finally, many of the 1970s conservation initiatives are still in place. We continue to have fuel-efficiency standards for automobiles and appliances. EP Act contains conservation provisions that embraced and carried forward energy conservation programs for commercial, residential, and industrial energy consumers.

ACKNOWLEDGMENTS

My colleague, Julia R. Richardson, contributed to the original paper on this subject, which appeared in *Natural Resources and Environment*, spring 1998.

REFERENCES

1. *Natural Gas Policy Act*, Public Law No. 95-621, 1978.
2. *Public Utility Regulatory Policies Act*, Public Law No. 95-617, 1978.
3. *Energy Tax Act*, Public Law No. 95-618, 1978.
4. *Powerplant and Industrial Fuel Use Act*, Public Law No. 95-620, 1978.
5. *National Energy Conservation Policy Act*, Public Law No. 95-619, 1978.
6. *Ad hoc Committee on Energy, National Energy Act*, 1977.
7. *National Energy Plan* P.52, Senate Committee on Energy and Natural Resources, Publication No. 95-16, May 1977.
8. Natural Gas Policy Act § 311, 2000.
9. *Ad hoc Committee on Energy*, National Energy Act, H.R. Rep. No. 543, 9th Cong., 1st Sess. 45, 1977.
10. *Order Revising Rules Requiring Clearance of Public Use Forms*, Order No. 528, 53 F.P.C. 1115, 1975.
11. *Regulation of Natural Gas Pipelines after Partial Wellhead Decontrol*, Order No. 436, FERC Statutes & Regulations, Regulations Preambles 1982–1985, 1985; 30,665.
12. *Pipeline Service Obligations and Revisions to Regulations Governing Self-Implementing Transportation, and Regulation of Natural Gas Pipelines after Wellhead Decontrol*, Order No. 636, FERC Statutes & Regulations, Regulations Preambles 1991–1996, 1992; 30,939.
13. *Energy Supply and Environmental Coordination Act of 1974*, Public Law No. 93-319, 1974.
14. *President Carter's Address on the National Energy Plan*, 1 Public Papers 657-58, 1977.
15. *National Appliance Energy Conservation Act*, Public Law No. 100-12, 1987.
16. *National Appliance Energy Conservation Act*, Public Law No. 100-357, 1988.
17. *The Energy Policy Act of 1992*, Public Law No. 102-486, 1992.

Natural Energy versus Additional Energy

Marc A. Rosen
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Abstract
Energy forms can be categorized in many ways, with one of the categories being natural or additional energy. Natural energy includes energy received directly and indirectly from the sun and energy derived from other natural forces while additional energy includes nonrenewable resources as well as energy forms that do not exist naturally but are produced. This article describes these energy categories and their applications. Also, energy forms, sources, and carriers are discussed and energy-conversion technologies are briefly described. Then, energy use, energy selection, and efficiencies are described, along with efficiency-improvement measures.

INTRODUCTION

Energy can exist in many forms and it can be converted from one form to another using energy-conversion technologies. Humans use energy carriers (often simply referred to as energy), which are produced from energy sources, in all aspects of living.

This article provides a basic understanding of natural energy and additional energy. Natural energy includes the energy received directly and indirectly from the sun as well as energy derived from other natural forces. Additional energy includes nonrenewable energy resources as well as energy forms that do not exist naturally but are produced by people. The main types of natural and additional energy have been studied extensively[1–7] and are listed in Tables 1 and 2, respectively.

From the breakdown of natural and additional energy in Tables 1 and 2, it is clear that energy resources are often categorized into two groups: (1) those generally acknowledged to be finite and nonrenewable and therefore not sustainable over the long term (e.g., fossil fuels, peat, uranium) and (2) those generally considered renewable and therefore sustainable over the relatively longer term (e.g., sunlight, wind, tides, falling water). Wastes (convertible to useful energy forms through, for example, waste-to-energy incineration facilities) and biomass fuels are also sometimes viewed as sustainable energy sources.

In this article, energy forms, sources, and carriers are explained. Natural and additional energy are discussed and energy-conversion technologies are briefly described. Then, energy use and factors in energy selection are discussed. Finally, efficiencies for energy use are presented, along with measures to improve energy efficiency.

Keywords: Natural energy; Additional energy; Renewable energy; Sustainable development; Efficiency; Energy source; Energy currency.

ENERGY FORMS, SOURCES, AND CARRIERS

Energy comes in a variety of forms, including fossil fuels (e.g., coal, oil, natural gas), fossil fuel-based products (e.g., gasoline, diesel fuel), uranium, electricity, work (e.g., mechanical energy in a rotating engine shaft), heat, heated substances (e.g., steam, hot air), and light and other electromagnetic radiation.

Energy sources (sometimes called primary energy forms) are found in the natural environment. Some are available in finite quantities (e.g., fossil fuels and fossil fuel-containing substances such as oil sands and peat). Some energy resources are renewable (or relatively renewable), including sunlight (or solar energy), falling water, wind, tides, geothermal heat, wood, and other biomass fuels (when the growth rate exceeds or meets the rate of use). Energy sources are often processed from their raw forms prior to use.

Energy carriers (sometimes called energy currencies) are the energy forms that we transport and use and they include some energy sources (e.g., fossil fuels) and processed (or secondary) energy forms (e.g., gasoline, electricity, work, heat). The processed energy forms are not found in the environment.

The distinction between energy carriers and sources is important. Energy carriers can exist in a variety of forms and can be converted from one form to another while energy sources are the original resource from which an energy carrier is produced. Sometimes misunderstanding between energy sources and carriers results because some energy sources are also energy carriers. For example, hydrogen is an energy carrier, not an energy source, and it can be produced from a wide range of resources using

Table 1 Types of natural energy

Direct solar radiation
Solar-related energy
Hydraulic energy (falling and running water, including large and small hydro)
Wave energy
Ocean thermal energy (from temperature difference between surface and deep waters of the ocean)
Wind energy
Biomass (where the rate of use does not exceed the rate of replenishment)
Nonsolar-related energy
Geothermal energy (internal heat of the earth)
Tidal energy (from gravitational forces of the sun and moon and the rotation of the earth)

Table 2 Types of additional energy

Energy sources
Fossil fuels
Conventional
Coal
Oil
Natural gas
Alternative
Oil shales
Tar sands
Peat
Nonfossil fuels
Uranium
Fusion material (e.g., deuterium)
Wastes (which can be used as energy forms or converted to more useful energy forms)
Energy currencies
Work
Electricity
Thermal energy
Heat (or a heated medium such as hot air, steam, exhaust gases)
Cold (or a cooled medium such as cold brine, ice)
Secondary chemical fuels
Conventional
Oil products (e.g., gasoline, diesel fuel, naphtha)
Synthetic gaseous fuels (e.g., from coal gasification)
Coal products (e.g., coke)
Nonconventional
Methanol
Ammonia
Hydrogen

various energy-conversion processes (e.g., water electrolysis, reforming of natural gas, coal gasification). Nevertheless, hydrogen is often erroneously referred to as an energy source, especially in discussions of its potential future role as a chemical energy carrier to replace fossil fuels.

NATURAL ENERGY

Natural energy includes the solar radiation incident on the earth and the energy forms that directly result from that radiation. Natural energy also includes the energy supplied by other natural forces such as gravitation and the rotation of the earth. The types of natural energy are summarized in Table 1. It is this energy that makes possible the existence of ecosystems, human civilizations, and life itself.

Solar Energy

Direct solar radiation is the main type of natural energy.[8] The daily energy output of the sun is 8.33×10^{25} kWh, of which the earth receives 4.14×10^{15} kWh. At any instant, the rate of solar energy reaching the earth is 1.75×10^{17} W, which is about 20,000 times greater than the total energy-use rate of the world. Solar radiation can be converted directly to electricity in photovoltaic devices. Also, solar energy can be collected as heat and used for thermal processes such as space and water heating or concentrated for use for in high-temperature heating and thermal electricity generation.

Most of the energy that enters the system of the earth and its atmosphere eventually exits to space. This concept can be demonstrated by considering the earth-sun energy balance (see Fig. 1). A general energy balance

$$\text{Energy input} - \text{Energy output} = \text{Energy accumulation}$$

can be applied to the earth when

- The energy input is the short-wave solar radiation entering the atmosphere.
- The energy output is the long-wave radiation exiting the atmosphere to space.
- The energy accumulation term is the increase in energy of the earth and its atmosphere.

The main implication of this global energy balance is that because the average temperature of the earth is

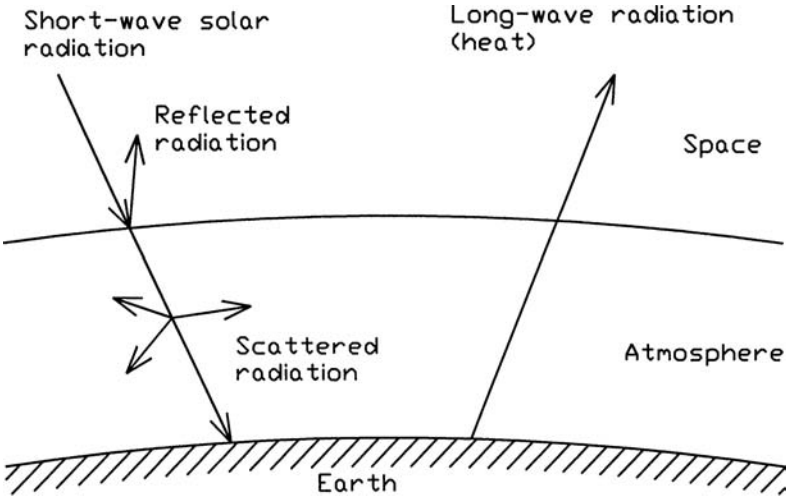


Fig. 1 Earth-sun energy balance.

relatively constant, the energy accumulation term is zero. Therefore, the energy output is equal to the energy input for the planet.

Global warming is caused by a disruption in the earth-sun energy balance. The main cause of global warming is increased releases of atmospheric "greenhouse gases" that absorb radiation in the 8–20 μm region. When greenhouse-gas concentrations increase in the atmosphere, energy output from the earth and its atmosphere (Fig. 1) is reduced while energy input remains constant. Thus, the energy accumulation term becomes positive, leading to an increase in the average temperature of the earth. Eventually, if concentrations of greenhouse gases in the atmosphere stabilize at new levels, the energy balance is re-established but at some higher average planetary temperature.

Solar-Related Energy

Several types of natural energy are a consequence of solar radiation. The most common is hydraulic energy, which includes falling and running water in natural settings such as rivers and waterfalls. Large-scale hydroelectric generating installations are common. Most economically utilizable hydraulic resources have already been developed. Recently, interest has grown in the potential uses of small-scale hydro.

Other forms of solar-derived energy are less common. Biomass energy includes wood and other forms of plants and organic matter. Biomass can act as a fuel itself or be converted into more desirable fuels. Several fast-growing trees have been identified as good candidates for biomass energy production. Biomass energy is only a renewable resource when the rate at which it is used does not exceed the rate at which it is replenished. Wind energy is used extensively in some countries (e.g., Denmark) for electricity generation, but it is not widespread.[9] Ocean thermal energy arises from the temperature difference between the surface and deep waters of the ocean. This temperature difference can be utilized to drive a heat engine, and several ocean thermal energy conversion (OTEC) devices have been tested. Wave energy systems have been proposed that take advantage of the motion of waves, although the potential contribution from wave energy is relatively small.

Nonsolar-Related Energy

The main types of natural energy in this category are geothermal energy, which exists as a consequence of the internal heat of the earth, and tidal energy, which is attributed to the gravitational forces of the sun and the moon and the rotation of the earth. Both of these energy sources have been used in limited ways.

ADDITIONAL ENERGY

Additional energy, often called secondary energy, includes both energy resources which are available in limited quantities and not renewable and energy forms produced by humankind.[2,7,10] Different types of additional energy are summarized in Table 2. Additional energy is related to the level of technological development of a society.

Two main categories exist for additional energy. The first consists of energy resources that are not renewable. The most common of these are fossil fuels, which are the basis of the economy for most industrialized countries. In addition to conventional fossil fuels, there exist alternative fossil fuels such as oil shales, tar sands, and peat. Other nonrenewable energy resources include uranium and fusion material (e.g., deuterium).

The second main category of additional energy is energy currencies that do not exist naturally. They include such basic energy forms as work, electricity, and thermal energy. The latter can include either heat or a heated medium like steam or cold water. Thermal energy in the form of heat or cold can be transported to users over long distances in district heating and cooling systems. District heating systems use centralized heating facilities to produce a heated medium which is transported to many users connected along a district heating network. For example, buildings in the cores of many cities are often connected by pipes through which hot water or steam flows to provide space and water heating. Similarly, district cooling involves the central production of a cold medium, which is transported to users through a piping network to provide cooling. Many cities and industrial parks utilize such district energy systems.

Wastes, which include recovered materials and energy that would otherwise be discarded, are also sometimes considered an energy source in the category of additional energy. Wastes can be used directly as energy forms or converted into more useful forms. Waste materials and waste heat can be recovered for utilization both within a facility and in other facilities where they are needed. For example, waste heat from hot gases (e.g., stack gases) and liquids (e.g., cooling-water discharges) can sometimes be recovered. Also, material wastes can be used in waste-to-energy incineration facilities, which burn garbage to provide heat and to generate electricity. Utilizing such wastes offsets the need for further supplies of external energy.

Additional energy also includes secondary chemical fuels. Some conventional examples include oil-derived products such as gasoline, diesel fuel, and naphtha, as well as synthetic gaseous fuels (e.g., from coal gasification) and coal products (e.g., coke). The types of nonconventional chemical fuels proposed are numerous and include methanol, ammonia, and hydrogen.

As many types of additional energy can be produced from energy resources or converted from other types of energy, it is important to understand and consider the entire life cycle of an energy product. The following life stages are usually included in assessments:

- Extraction or collection of raw energy resources
- Manufacturing and processing of the desired energy form(s)
- Transportation and distribution of the energy to users
- Energy storage
- Use of the energy to provide services and tasks
- Recovery and reuse of output energy that would otherwise be wasted (e.g., waste heat recovery)
- Recycling of wastes from any of the above steps
- Disposal of final wastes (e.g., materials such as stack gases and solid wastes including ash)

For example, the life cycle of a general energy form may involve the following chain of events:

$$\text{Raw resource} \rightarrow \text{Finished resource} \\ \rightarrow \text{Energy product} \rightarrow \text{Waste} \rightarrow \text{Waste disposal}$$

Two or more types of additional energy can be simultaneously produced in some systems. For example, cogeneration is a process that usually refers to the combined generation of electricity (or work) and heat (or a heated medium). Trigeneration refers to an extended cogeneration process in which cooling is provided as a third product.

Some examples of the different life cycles for electricity generation methods from a range of energy sources are presented in Table 3. In that table, methods based on fossil fuels and nonfossil resources are considered. In addition, electricity generation from different energy sources via a less conventional technology, fuel cells, is considered in Table 4. All processes in Tables 3 and 4 include as a final step in the life cycle equipment disposal, although that step is not shown.

ENERGY USE

Energy use involves the production of useful energy forms through energy-conversion processes as well as energy transportation, distribution, and storage, and the utilization of energy resources and processed forms of energy to provide services and perform tasks.

Energy-Conversion Technologies

Desired energy carriers are produced from energy sources or converted from one form to another using energy-conversion technology. The energy-conversion technology appropriate in a given instance depends on the initial (or source) and final (or desired) energy forms. Conventional energy-conversion technologies include hydroelectric, fossil fuel, and nuclear generating stations; oil refineries; engines and motors; and heaters. Some examples of less conventional energy-conversion technologies include fuel cells, solar photovoltaics, and high-efficiency or clean technologies for fossil fuels (e.g., combined-cycle systems).[4] Some energy-conversion technologies yield more than one product (e.g., cogeneration). Various technologies for electricity generation are listed in Tables 3 and 4.

A major energy conversion process is combustion, which drives furnaces, engines, transportation vehicles, etc., and which can be expressed for a hydrocarbon fuel as

Table 3 Life cycles for selected energy sources and methods for electricity generation

Method and energy source	Energy-conversion processes in the life cycle
Fossil fuel-based methods	
Natural gas	Extraction via gas well → transport → processing → natural gas-fired power plant
Coal	Coal mining → transport → processing → coal-fired power plant
Oil	Extraction via oil well → transport (tanker or pipeline) → refining → oil-fired power plant
Nonfossil-based methods	
Nuclear	Uranium mining → transport → processing → nuclear power plant
Hydro	Hydraulic turbine
Wind	Wind generator
Solar (thermal)	Solar energy collection → thermal power plant
Solar (non-thermal)	Solar photovoltaic panels
Geothermal	Well to geothermal source → geothermal power plant
Ocean (thermal)	Ocean thermal energy conversion (OTEC) power plant

$$\text{Hydrocarbon fuel} + \text{Air(mainly } O_2 \text{ and } N_2) \rightarrow CO_2 + H_2O + N_2 + \text{Other substances}$$

The other substances are leftover reactants and other reaction products. The reaction for stoichiometric combustion (i.e., complete combustion with a balanced amount of reactants and no reactions that yield other products) of a general hydrocarbon (C_mH_n) in air (treated as only nitrogen and oxygen) can be written as

$$C_mH_n + (m + n/4)(O_2 + 3.76N_2) \rightarrow mCO_2 + n/2H_2O + 3.76(m + n/4)N_2$$

Here, m and n are variables that take on different values for different hydrocarbons (e.g., in approximate terms, $m=1$ and $n=4$ for natural gas, which is mainly methane, $m=1$ and $n=2$ for oil, $m=1$ and $n=1$ for typical coal, and $m=1$ and $n=0$ for pure carbon). Several important points stem from these expressions. First, carbon dioxide is an inherent product of the combustion of any carbon-containing fuel. Generally, the only way to avoid carbon dioxide emissions is to eliminate the use of carbon-based fuels (e.g., by using hydrogen as a fuel). Second, if insufficient oxygen is available (or if mixing is poor or reaction time is short), the carbon will often only partially combust, yielding CO. Third, excess air (over stoichiometric) is usually used to improve fuel burn-up, often lowering combustor efficiencies because some of the fuel energy must go towards heating the excess air. Fourth, the fuel is often not a pure hydrocarbon and contains other substances, e.g., sulphur. Sulphur combusts yielding heat and reacts to sulphur dioxide.

Energy Selection

Energy selection involves choosing both energy carriers and sources. The selection often depends on the energy service to be provided and the energy-conversion technologies available. Some energy selections involve energy or fuel substitution (e.g., heating with natural gas rather than electricity).

To prevent environmental impact, energy sources that are renewable and cause relatively lower environmental impacts are usually preferred.[11–13] By extension, preference is also placed on energy sources and carriers that can be used with higher efficiency and more environmentally benign energy-conversion technologies (e.g., boilers having low emissions of nitrogen oxides, NO_x).[14]

Renewable energy sources fall into the category of natural energy (Table 1). They usually are derived from sunlight or solar-derived sources (e.g., wind, waves) and

Table 4 Life cycles for electricity generation from various energy sources using fuel cells

Energy source	Energy-conversion processes in the life cycle
Fossil fuels	Fossil-based hydrogen production → fuel cell
Nonfossil energy sources	Hydrogen production → fuel cell
High-temperature heat (e.g., solar)	Hydrogen production via thermochemical cycle → fuel cell

are sustainable. A major barrier to renewable energy sources is that they are usually more costly to use than nonrenewable energy sources such as fossil fuels, although this observation is not true in some niche applications.

The use of fossil fuels leads to combustion emissions. Problematic pollutants are generally lower for fuels having higher hydrogen-to-carbon atomic ratios, so natural gas is more benign than oil, which is more benign than coal. There are no emissions during normal operation of nuclear power facilities except for the spent fuel, which remains radioactive for many years. Although falling and running water derived from solar energy drives hydroelectric generation, flooding of lands sometimes occurs. The use of renewable energy is relatively benign, but resources are required to build the relevant energy-conversion technologies and large land tracts are often needed (e.g., for solar collectors or wind generators).

Energy Efficiency and Other Measures of Merit for Energy Use

Several efficiencies can be defined for energy use. Other measures are also often used to assess the merit of energy use relative to other criteria. One important measure is energy intensity, which reflects the energy use per some unit of output, e.g., energy consumption per monetary unit of Gross Domestic Product for a country or region. Because they deal with raw materials, primary processing industries (e.g., petroleum/coal production, chemicals production, primary metals) generally have higher energy intensities than intermediate and secondary processing industries which, although they come in a wide variety of forms, tend to deal with more finished products.

Numerous methods and technologies exist for improving energy efficiency:

Use of High-Efficiency Devices

Energy efficiency can generally be increased via high-efficiency versions of energy-intensive devices (e.g., home appliances, furnaces, air conditioners, motors, boilers). New lighting fixtures (including bulbs, reflectors, diffusers, and ballasts), for instance, have significantly higher efficiencies and longer lives than older equipment. Lighting efficiency can also be increased by task lighting (i.e., directing light where it is needed), reducing lighting to levels adequate for the human eye and the nature of the facility, and using of timers, dimmers, and occupancy sensors.

Energy Leak and Loss Prevention

Energy losses are associated with energy leaks. Approaches to avoid such energy leaks include (1) applying leak-prevention technologies and methods (e.g., sealing leaks in storage vessels and pipelines and insulating to reduce undesired heat flows) and (2) inspecting periodically to detect leaks and initiate appropriate actions.

Application of Advanced and Integrated Energy Systems

Many advanced energy systems feature increased efficiencies compared to conventional technologies. Efficiency also can be increased by advantageously integrating energy systems so waste becomes input to other processes. Waste-recovery and waste-to-energy plants normally increase efficiency by offsetting some of the need for external energy. Cogeneration and trigeneration are often more efficient than separate processes for the individual products. Compared to the alternative of each facility having its own heating and cooling plant, district heating and cooling often provide increased efficiency and reduced environmental impact because many efficiency-improvement and environmental-control measures are possible in centralized, large-scale facilities.

Improved Monitoring, Control, and Maintenance

Energy efficiency can be improved through better monitoring, control, and maintenance of operations so that performance degradation from design specifications is avoided. Computer systems allow automated acquisition of frequent and widespread readings and diagnostic evaluations. The life span of energy equipment can be increased through diligent maintenance (e.g., cleaning, lubricating, calibrating, tuning, regular testing, periodic overhauling, and replacing consumable items regularly).

Improved Matching of Energy Supplies and Demands

Rather than supply an energy form of a level that greatly exceeds that required for a specific energy demand, it is usually more efficient to supply an energy form of a better matched level. For instance, better matching can often be achieved for heat-transfer flow temperatures (e.g., for space heating at about 22°C, a heat supply of perhaps 40°C could be sufficient, rather than furnace combustion gases at hundreds of degrees). Thus, industrial waste heat and low-temperature cogenerated heat can satisfy some heating needs.

Energy Storage

Sometimes available energy supply exceeds demand and the energy is not utilized. At times, the supply is not controllable (e.g., sunlight may be available for heating during the day whether or not it is needed), while in other cases the supply may not be easily alterable (e.g., the waste heat from a factory). Energy storage can improve system efficiency by storing energy between times when it is available and when it is needed.

Improved Building Envelopes

The energy efficiency of a building can be improved by increasing insulation to reduce heat infiltration in summer and heat loss in winter, applying weather stripping and caulking to reduce air leakages, and utilizing advanced and high-efficiency windows. The latter reduce heat losses by using multiple glazings that are sometimes separated by insulating gases or a vacuum, electronic and photosensitive windows which automatically reflect or absorb excessive sunlight, low-emissivity window coatings that increase a window's resistance to heat loss, and efficient window shades that adjust themselves using photosensitive sensors to block excessive sunlight while permitting natural room lighting.

Use of Passive Strategies to Reduce Energy Use

Passive, as opposed to active, methods can be used to reduce energy use. Some examples: adjusting design temperatures such as those used for space heating in winter and space cooling in summer; shutting off energy devices during periods of nonuse and using timers to control operating hours; using daylight harvesting to reduce artificial lighting needs; applying "free cooling" by using cool outdoor air when it is available to cool warm indoor spaces rather than using active air conditioning; using solar radiation to heat buildings by exploiting the thermal storage capacity of buildings; and locating trees, windows, and window shades so as to keep buildings cool during summers.

Use of Exergy Analysis

Exergy analysis is used to improve and optimize the efficiency and performance of energy and other systems and processes.[1,10,15,16]

EXAMPLE

Many examples can illustrate the methods and technologies described in "Energy Efficiency and Other Measures of Merit for Energy Use" for improving energy efficiency. Here, we focus on the use of high-efficiency lighting. New fixtures have significantly higher efficiencies and longer lives than older equipment. For example, the efficiencies for some light sources (in lumens of light delivered per watt of electricity consumed) are: incandescent (10–30), mercury (20–55), fluorescent (20–60), high-pressure sodium (50–130), and low-pressure sodium (80–155). Thus, the same amount of light can be delivered with high-efficiency lighting using less than 10% of the electricity required for an incandescent bulb.

On a regional or national level, the savings can be significant. Because lighting is a significant energy consumer, accounting for about 20% of U.S. electrical energy use, the use of high-efficiency lighting across the United States could reduce that country's electricity consumption by 18% (i.e., 90% of the 20% of electricity use attributable to lighting). Of course, cost and other considerations limit the implementation of high-efficiency lighting.

Note that other measures to increase lighting efficiency could further reduce the electricity used for lighting. If, for instance, electricity use for lighting is reduced by an additional 5% through the application of (1) timers, dimmers, and occupancy sensors to turn lights off when rooms are unoccupied, (2) reduced lighting intensities, and (3) task lighting, then electricity use in the United States could be reduced by 19% (i.e., 95% of the 20% of electricity use attributable to lighting).

CONCLUSION

Categorizing energy forms as natural and additional energy allows for an improved understanding of energy systems. Natural and additional energy are important in societies, particularly because they affect many facets of life and living standards. This article describes these energy categories and their applications. Also discussed here are energy forms, sources, carriers, and technologies; energy use; energy selection and efficiencies; and efficiency-improvement measures.

ACKNOWLEDGMENTS

The author is grateful for the support provided for this work by the Natural Sciences and Engineering Research Council of Canada.

Glossary

m: number of carbon atoms in a hydrocarbon
n: number of hydrogen atoms in a hydrocarbon

REFERENCES

1. Rosen, M.A.; Scott, D.S. Entropy production and exergy destruction—Part II: Illustrative technologies. Int. J. Hydrogen Energy **2003**, *28* (12), 1315–1323.
2. Organization for Economic Cooperation and Development *Energy: The Next Fifty Years*; OECD: Washington, DC, 1999.

3. Organization for Economic Cooperation and Development *World Energy Outlook*; OECD: Washington, DC, 1998.
4. *Handbook of Energy Systems Engineering:Production and Utilization*; Wilbur, L.C. Ed.; Wiley: Toronto, 1985.
5. Niele, F. *Energy: Engine of Evolution*; Elsevier: Oxford, U.K., 2005.
6. O'Callaghan, P.W. *Energy Management*; McGraw-Hill: New York, 1993.
7. Goldemberg, J.; Johansson, T.B.; Reddy, A.K.N.; Williams, R.H. *Energy for a Sustainable World*; Wiley: New York, 1988.
8. Wright, S.E.; Rosen, M.A. Exergetic efficiencies and the exergy content of terrestrial solar radiation. Trans. ASME J. Solar Energy Eng. **2004**, *126* (1), 673–676.
9. Sahin, A D.; Dincer, I.; Rosen, M A. Thermodynamic analysis of wind energy. Int. J. Energy Res. **2006**, *30* (8), 553–566.
10. Rosen, M.A.; Scott, D.S. Entropy production and exergy destruction—part I: Hierarchy of earth's major constituencies. Int. J. Hydrogen Energy **2003**, *28* (12), 1307–1313.
11. *The Wiley Encyclopedia of Energy and the Environment*; Bisio, A.; Boots, S. Eds.; Wiley-Interscience: Toronto, 1996.
12. Liu, P.I. *Introduction to Energy and the Environment*; Van Nostrand Reinhold: New York, 1993.
13. Winteringham, F.P.W. *Energy Use and the Environment*; Lewis Publications: Boca Raton, Florida, 1992.
14. Claridge, D.; O'Neal, D.; Roberts, M. Improving energy conservation retrofits with measured savings. ASHRAE J. **1991**, 14–22.
15. Kotas, T.J. *The Exergy Method of Thermal Plant Analysis*; Krieger: Melbourne, FL, 1995; [reprint edition].
16. Rosan, M.A. Exergy: Analysis. In *Encyclopedia of Energy Engineering and Technology*, Capehart, B., Ed.; Taylor & Francis: New York, 2007.

Net Metering

Steven Ferrey
Suffolk University Law School, Boston, Massachusetts, U.S.A.

Abstract
Eighty percent of states have adopted net metering, a regulatory innovation to implement decentralized renewable power alternatives. Net metering provides the most significant boost of any policy tool at any level of government—both qualitatively and quantitatively—to decentralize American power sources. Forty states to date are implementing net metering. Net metering, legally, if not physically, runs the retail utility meter backwards when a decentralized or renewable energy generator moves power back to the grid. With the demand for electricity increasing in both developed and developing nations, whether a new power supply is developed in a centralized or decentralized mode has profound implications. The manner in which states encourage or discourage the creation of decentralized dispersed energy sources through various regulatory, subsidy, and metering initiatives will sculpt electric energy's future.

INTRODUCTION

By turning the meter backwards, net metering effectively compensates the generator at the full retail rate for transferring just the wholesale energy commodity. While most states compensate the generator for excess generation at the avoided cost or market-determined wholesale rate, as the table below shows, some states compensate the wholesale energy seller for the excess at the fully loaded and much higher retail rate.

Net metering can pay the eligible renewable energy source as much as four times more for power than the amount paid to any other independent power generators, and much more than the time-dependent value of this power to the purchasing utility. A 400% price advantage over the competition in these 40 states provides a nationwide platform to support decentralized energy production.

Electricity is a unique energy form—it cannot be stored or conserved with any efficiency. Therefore, electricity has substantially different value at different hours of the day, different seasons of the year, and different places in the utility system. Contrary to this physical reality, net metering and billing treats all power at all hours as being tangibly storable or bankable and as if it has equal value when, in fact, it does not.

By ignoring interim actual physical transfers of power occurring at all minutes and hours of the month and by recognizing only the net balance of the transactions at the end of the month or quarter, net metering assumes all electricity generated and transmitted has equal average value. This is not accurate at the wholesale level, it is not the case with power trading, and it is not the case in those 18 states where retail competition has been promoted with deregulated competitive retail markets. In deregulated states, wholesale power is differentially valued and priced each hour of each day of the year. The market and regulatory reality is contrary to the key implicit premise of the eventual U. S. Federal Energy Regulatory Commission (FERC) decision in the MidAmerican case.

It is even possible to "game" the system with net metering by selling power to the utility at the netted average retail price in off-peak late evening hours when the customer/generator has no need for the power and the utility has surplus power. Other utility ratepayers will ultimately be left to make up the resulting revenue deficit.

LEGAL AND POLICY CONTROVERSY

The Action in Iowa

Although it originated in Minnesota, net metering received its initially controversial legal validation in a case that occurred in Iowa. Thus, both in its inception and in its later legal validation, net metering is a product of the nation's heartland.

The Iowa Utility Board (IUB) originally approved net metering. On August 24, 1999, the Polk County district court of Iowa issued a decision that would eventually evolve to impact energy policy throughout the United States. The district court ruled that federal law preempts Iowa's regulatory authority, which is used to compel a utility to permit small generating facilities—such as on-site dispersed wind and solar facilities—to interconnect with the power grid under net billing or metering arrangements. The Polk County district court ruled that small electric generation facilities are Qualified Facilities

Keywords: Utility rates; Metering; Net billing; Net sales; Renewable energy sales; Renewable energy policy.

(QFs) governed by the federal Public Utility Regulatory Policies Act of 1978 (PURPA), which precludes sales of excess power generated by QFs at rates in excess of the purchasing utility's avoided cost.

FERC defines avoided costs as "the incremental costs to an electric utility of electric energy or capacity or both which, but for the purchase from the QF or QFs, such utility would generate itself or purchase from another source." The Iowa court also ruled that if these small generating facilities are not QFs under PURPA, then they are public utilities engaged in the wholesale sale of power in interstate commerce and are therefore governed by the Federal Power Act and regulated by FERC. In either event, federal law governed all activities.

Central to the court's conclusion is its determination of what net billing involves at its core—a sale of electricity. From this finding, it followed that irrespective of the volume of power involved, the transactions are considered wholesale sales; as soon as the energy flowing to the utility is commingled with other energy in the power grid, it is sold in interstate commerce. FERC has exclusive jurisdiction to set rates governing wholesale sale of electricity in interstate commerce. Accordingly, FERC's exclusive rate-setting authority over interstate commerce under the Federal Power Act preempts the IUB's ruling.

The Federal Action at FERC

When the IUB sought an appeal of the district court opinion to the Iowa Supreme Court, the utility, Midamerican, moved laterally and took its grievance to FERC, which accepted jurisdiction. In its decision, FERC held that the IUB decisions were not preempted by federal law. FERC held that no sale occurs when an individual homeowner installs generation and accounts for its dealings with the utility through the practice of net metering.

The court ruled that a change of title to power does not constitute a sale. FERC ignored the physical reality of the transfer of the electrons and left regulation of the netting aspect of the transaction to the states. In the MidAmerican case, FERC stated that "there is no sale (for end use or otherwise) between two different parties when one party is using its own generating resources for the purpose of self-supply of station power, and accounting for such usage through the practice of netting."

According to FERC, the energy flowing from a generator to a utility under net metering is an exchange or offset rather than a sale. Proponents of the FERC exchange view argue that there has been no value assigned to the energy transmitted because the single meter is read only at the end of the billing period. Because an assignment of value cannot take place prior to the meter reading, the criterion establishing a sale will not be met. Therefore, only an exchange has taken place, yielding no federal PURPA requirements.

Proponents of the contrary view note that even if no payment is yet due, as the electrons pass through the meter their title passes, a monetary obligation occurs and is due in the future under either a power purchase agreement or the utility tariff. With net metering, there is no doubt that power is physically transferred to the local utility, often at times when it does not wish to take that power and it cannot sell that power to consumers, and it must ground or dispose of such power. An actual physical transfer or exchange of generated electrons and electric current to the utility occurs at the discretion of the generator. The utility takes the title, pays for these electrons (or nets them against others sold in the opposite direction to the generator), comingles these electrons in its system, and resells them to other third parties or disposes of them as it can.

Both a physical and legal transfer has occurred, with all the attributes of a legal transfer. Given that a contract or tariff amount is eventually required to be paid, netted, or bartered by the utility, this would seem to resemble a sale. There is also no doubt that the transaction in dispute occurs at the wholesale level, although it is netted against a retail sale.

In essence, FERC held that it was within state authority to only recognize a transaction as not what occurs on an instantaneous, hourly, daily, or even weekly physical basis. Rather, it ruled that it is within state authority to recognize a transaction through single or dual meters once per billing period (monthly or quarterly)—the net result of the transfer of power to and from a distributed generator and a local utility—as a positive or negative retail sale. In other words, it is legally permissible to take a periodic (retail) snapshot of the net transactions between these two entities and treat these as the only point of sale. FERC allows states to make the netting decision, which, at its core, is a determination to recharacterize instantaneous (FERC jurisdictional) wholesale sales as (state jurisdictional) retail sales.

WHAT THE STATES HAVE IMPLEMENTED

While Minnesota was the first state to enact net metering, 29 other states adopted some form of net metering between 1980 and 2000. Since the 2001 FERC decision in MidAmerican sanctioned net metering, additional states have implemented net metering.

The table below sets forth the types of technologies eligible for net metering in representative states, the types of eligible participating customers, size limits, and what is done with the credit, if any, earned by the customer. Notice that while most states include renewable energy technologies, there is significant variation. Some states do not include hydroelectric technologies in their definitions because of environmental group opposition to additional river damming for the purpose of generating hydroelectric power. Some states also do not include municipal solid waste (MSW) trash-to-energy technologies as eligible because of an objection to the burning of municipal trash (as opposed to landfilling or recycling trash). A few states with significant in-state coal industries initially defined coal as an eligible renewable technology. In short, the manner in which states define eligible technologies is subject to state policy discretion.

States also vary greatly with regard to how large eligible installations are. While most states limit the size to smaller installations scaled to on-site use, a few allow quite large capacity facilities to net meter their power. Similarly, some states limit eligible participants to residential customers of the utility, or residential and commercial customers, while others extend participation to all (including commercial and industrial) customers. Commercial and industrial customers are some of the most cost-effective users of on-site cogeneration technologies.

How states treat net energy generation (NEG) is one of the more controversial aspects of net metering. NEG is the net surplus of electricity sold to the utility compared to electricity purchased from the utility over a given (typically monthly) billing period. Some states allow any such surplus to be carried over as a credit against the next month, with some limiting the duration of this carry-over to a year. At the end of the year, the surplus is either forfeited to the utility or to low-income energy assistance programs administered by the utility (which effectively pay the utility bill of customers who have not paid). Still, other programs allow the customer to receive cash for the NEG.

Each of the states operates as its own net metering laboratory, with various types of eligible technologies, participating customers, application sizes, and ultimately, NEG credit systems. This provides several models for implementing net metering. Collectively, net metering provides the single greatest policy incentive for on-site distributed generation in the United States. The table below illustrates many of the states' net metering programs and distinctions.

State	Eligible technology	Eligible customers limits	Size limits	Price
Arizona	Renewables and cogeneration	All customer classes	$\leq$100 kW ($\leq$10 kW for AZ Public Service customers)	NEG* carried forward; granted to utility at year end
Arkansas	Most renewables, fuel cells, and microturbines	All customer classes	$\leq$25 kW residential $\leq$100 kW commercial	Monthly NEG granted to utilities
California	PV, landfill gas, anaerobic digestion, solar, and wind	All customer classes	$\leq$1000 kW	Annual NEG granted to utilities
Colorado	Small hydro, wind, and PV	Varies by utility	Varies Xcel: $\leq$10 kW	NEG carried forward month-to-month
Connecticut	Renewables, cogeneration, MSW, and fuel cells	Residential customers	$\leq$50 kW fossil tech $\leq$100 kW renewables	NEG purchased at spot market price
Delaware	Solar, wind, and other renewables	Residential and commercial	$\leq$25 kW	Not specified
District of Columbia	Solar, PV, wind, biomass, fuel cells, and microturbines	All customer classes	$\leq$100 kW	Not specified
Florida	PV	All customer classes	Not specified	Customers receive full retail credit
Georgia	PV, wind, fuel cells	Residential and commercial	$\leq$10 kW residential $\leq$100 kW commercial	Monthly NEG carried forward; granted to utility at year end
Hawaii	PV, wind, biomass, and hydro	Residential and small commercial	$\leq$10 kW	Monthly NEG granted to utilities
Idaho	Renewables and fuel cells	All customer classes	Varies by utility	Monthly NEG purchased at retail rate
Illinois	PV, biomass, and wind	Retail customer classes; Commonwealth Edison only	$\leq$40 kW	NEG purchased at avoided cost monthly plus annual payment to bring payment to retail rate

(Continued)

Nat–Perf

(Continued)

State	Eligible technology	Eligible customers limits	Size limits	Price
Indiana	Renewables and cogeneration	All customer classes	≤1000 kWh/month	Monthly NEG granted to utilities
Iowa	Renewables and MSW	All customer classes	No limit per system	Monthly NEG purchased at avoided cost
Kentucky	PV, hydro, and wind	All customer classes	≤10 kW	NED carried to next month
Louisiana	Renewables, fuel cells, and microturbines	All customer classes	Residential ≤25 kw; Commercial and Agriculture ≤100	Not specified
Maine	MSW, cogeneration, renewables, and fuel cells	All customer classes	≤100 kW	NEG carried forward; annual NEG granted to utilities
Maryland	Solar and PV	Residential and schools only	≤80 kW	TBD by PSC
Massachusetts	MSW, renewables, cogeneration, and fuel cells	All customer classes	≤60 kW	Monthly NEG purchased at avoided cost
Minnesota	Renewables, MSW, and cogeneration	All customer classes	≤40 kW	NEG purchased at utility average retail energy rate
Montana	Fuel cells, geothermal, solar, wind, and hydro	All customer classes	≤50 kW	Carried forward to next month; annual NEG granted to utility
Nevada	Wind, hydro, biomass, solar, and wind	All customer classes	≤30 kW	Granted to utility
New Hampshire	PV, wind, and hydro	All customer classes	≤25 kW	NEG credited to next month
New Jersey	PV and wind	Residential and small commercial	≤100 kW	Carried forward to next month; annualized NEG purchased at avoided cost
New Mexico	Renewables, microturbines and cogeneration	All customer classes	≤10 kW	NEG credited to next month, or monthly NEG purchased at avoided cost (utility choice)
New York	PV only (biogas for farms)	Residential and agriculture only	≤10 kW (PV); ≤400 kW (biogas)	NEG credited to next month; annualized NEG purchased at avoided cost
North Dakota	Renewables, MSW, and cogeneration	All customer classes	≤100 kW	Monthly NEG purchased at avoided cost
Ohio	Renewables, microturbines, and fuel cells	All customer classes	No size limit (≤100 kW for microturbines)	NEG purchased at unbundled generation rate
Oklahoma	Renewables, MSW, and cogeneration	All customer classes	≤100 kW and ≤25,000 kWh/year	Monthly NEG granted to utility
Oregon	Solar thermal, wind, fuel cells, and hydro	All customer classes	≤25 kW	Annual Neg granted to low-income programs, credited to customer, or other use determined by commission
Rhode Island	MSW, cogeneration, renewables, and fuel cells	All customer classes	≤25 kW	Carried forward month-to-month; annual NEG granted to utilities

(Continued)

Nat–Perf

(Continued)

State	Eligible technology	Eligible customers limits	Size limits	Price
Texas	Renewables only	All customer classes	≤50 kW	Monthly NEG purchased at avoided cost
Utah	Solar, thermal, PV, wind, hydro, and fuel cells	All customer classes	≤25 kW	NEG credited to next month; any unused credit granted to utility at end of calendar year
Vermont	PV, wind, fuel cells using renewable fuels, and anaerobic digesters	Residential, commercial and agricultural	≤15 kW; farm giogas ≤150 kW	NEG credited to the following month; annual NEG granted to utilities
Virginia	Solar thermal, wind, PV, and hydro	Residential and commercial	≤10 kW residential ≤25 kW nonresidential	NEG carried forward indefinitely
Washington	Solar, wind, fuel cells, and hydro	All customer classes	≤25 kW	NEG carried forward monthly; annual NEG granted to utility
Wisconsin	Renewables, MSW and cogeneration	All customer classes	≤20 kW	Monthly NEG purchased at retail rate for renewables, avoided cost for nonrenewables
Wyoming	PV, wind, biomass and hydro	All customer classes	≤25 kW	NEG carried forward monthly; Annual NEG purchased at avoided cost
Puerto Rico	Renewables	Residential	≤50 kW	NEG carried over month-to-month; unused credits at end of year purchased at avoided cost

Notes: KW, kilowatt, a measure of electric generation capacity; MSW, municipal solid waste-to-electricity conversion, usually accomplished by combusting the waste in a boiler to drive a generator; NEG, net electric generation, the surplus of electricity sold to the utility over the amount of electricity purchased from the utility over a given period (month or year); PV, solar photovoltaic technology.

CONCLUSION

States are allowed to utilize largely invisible transactions through the rate base to subsidize projects through net metering without FERC oversight. Net metering and billing policy constitute the most important of four policy supports for the renewable energy industry initiatives in the United States. In addition to net metering and billing, almost two dozen states have elected to establish renewable portfolio standards and/or system benefit charges that support renewable energy trust funds. However, among these state-level renewable energy policy initiatives, net metering and billing is not only the most pervasive, having been adopted in 40 states, but it also provides the most significant financial advantage to generators.

BIBLIOGRAPHY

1. Some of the information in this chapter previously appeared in S. Ferrey's "Nothing But Net", 14 Duke Environmental Law & Policy Forum 1(2003).

Nuclear Energy: Economics

James G. Hewlett
Energy Information Administration, U.S. Department of Energy, Washington, D.C., U.S.A.

Abstract
This article attempts to answer the question: "Is Nuclear Power Economic?" There are two major factors influencing the economics of nuclear power, and since major uncertainties with both of them exist, it is impossible to give an unqualified answer. The first factor is the cost of building the nuclear power plant. Analysis has found that costs must fall from about $2000 per kW of capacity to about $1500 per kW before nuclear power would be economic. The second factor is the environmental impact of the alternative. The economics of nuclear power would be greatly improved if all of the external costs related to global warming were included in the cost of generating electricity from fossil fuel-fired power plants. Nuclear power plants have their own set of environmental costs, but since they are incurred over hundreds of years, their "present cost today" is very small. Some have objected to discounting expenses incurred over very long time periods because this procedure represents a strong incentive to impose large costs on future generations. However, such equity considerations are outside the realm of economic analysis.

INTRODUCTION

The objective of this article is to answer the question: "Is Nuclear Power Economic?" The discussion will focus on the two most important factors surrounding the economics of nuclear energy. The first and most obvious factor is the cost of constructing the nuclear power plant and the second involves a complex set of environmental considerations. As will be seen here, there are major empirical uncertainties in estimating nuclear capital costs. Additionally, because some of the environmental costs associated with nuclear power will be borne by future generations, there are some ethical considerations in evaluating nuclear power that are outside the realm of economic analysis. Thus, there is no "clear cut" answer to the question of whether nuclear power is economic.

Moreover, in its most general form, this question is too broad to analyze in a short article, and therefore, in three areas the focus will be relatively narrow. First, that the economics of nuclear power is a relative concept comes from the fact that this technology comprises just one way of generating electricity. Nuclear power will, therefore, be compared to large coal and combined-cycle, natural gas-fired power plants. Secondly, the article will focus on the United States, because large amounts of data and fossil fuel price projections are in the public domain. However, the analysis, as opposed to the conclusions, is sufficiently broad, so that it could be used to examine the economics of nuclear power in other countries. Third, the economics of nuclear power have a temporal element since costs and prices change over time. Currently, large amounts of excess capacity exist in the United States, and new baseload capacity will probably not be needed until about 2010–2015. Additionally, because of the complex licensing process and relatively long construction lead times, the earliest a nuclear power plant could become operational is about 2010–2012. Thus, the economics of nuclear power as of around 2010 will be analyzed.

Since the focus will be on the near term, technological improvements to existing technologies will be limited. As will be seen shortly, global warming considerations are one of the major factors affecting the economics of nuclear power. A number of fossil fuel carbon sequestration technologies are under development, but since none of them will be commercially available until about 2020, they will not be considered here. Moreover, the analysis will be prospective in the sense that the focus is on plants to be built and operated in the future. Since there are "no facts about the future," this type of prospective analysis requires numerous assumptions about what costs will be like 5–45 years from now. Unfortunately, because of length considerations, most of them will not be discussed here. Instead, the reader will be referred to other documents.

NUCLEAR CAPITAL COSTS

Construction costs represent about 70%–80% of the total cost of generating electricity from nuclear power plants, and therefore, the estimates of nuclear capital costs are one of the most important factors affecting the economics of this technology. There are two general ways of estimating the cost of building a nuclear power plant, or any other capital project. One common approach is to derive an engineering based "bottom-up" estimate. That is, the cost of every small component

Keywords: Nuclear power economics; Nuclear power construction costs; Nuclear decommissioning; Nuclear waste.

(e.g., a pump) and the associated labor installation expenses would first be estimated, with the total derived by simply adding all the component costs. Factors such as bad weather or labor strikes almost always occur but are typically not incorporated in the bottom-up estimates. To account for this, a contingency allowance is generally included. (The contingency is typically a fixed percent of the baseline cost.) This approach has been used for many years and, many times, will provide accurate estimates for plants that have been built. However, studies have shown that "bottom-up" estimates for plants with new designs that have not been built before generally tend to be too low.[1]

The second approach is to derive the estimate by using aggregate realized capital costs of similar plants. When this is done, the design of the plant in question should be similar to the ones whose costs are being used to derive the estimate. Many times, statistical techniques are used to standardize the plants, which generally poses a problem. As will be seen shortly, additional problems exist when realized costs in foreign countries are used.

Looking first at realized costs, the capital cost of recently completed light-water reactors abroad with advanced designs appears to be about $1700–$2000 per kilowatt (kW) of capacity. In particular, in the late 1990s, General Electric (GE) completed the design of an Advanced Boiling Water Reactor (ABWR) with improved and simplified safety features. A few of these units have been built in Japan and, according to a recent MIT study, their construction costs were about $1800–$2000 per kW.[2] Additionally, Korea used the design of the System 80+ as the basis for their advanced Pressurized Water Reactors (PWRs). Again, according to the MIT researchers, the capital cost of these advanced PWRs was about $1800 per kW. Lastly, a utility in Finland recently decided to build an advanced PWR designed by a group of French and German firms. Construction of the plant will begin in 2005 or 2006, and the reported capital cost was about $1700 per kW. If this foreign experience is at all indicative of what would occur in the United States, the cost of building new units here with advanced designs that have been built somewhere in the world is about $1700–$2000 per kW.

It must be stressed that extrapolating the cost of reactors built in one country to another must be done with great care. First, there are problems with using official exchange rates to convert costs that are expressed in one currency to another when exchange rates are volatile, or when they are distorted by government intervention (an example here is Japan that has intervened in world currency markets to support the yen. This was done to improve the competitive position of Japanese exports). Thus, all the costs cited above were converted to U.S. dollars using the Purchasing Power Parity Index. This index is basically an estimate by the Organization for Economic Cooperation and Development (OECD) of what exchange rates should be in a perfect world without volatile or distorted currency markets.

Second, there is little information about the items that are included in the costs. Typically, a utility will make contractual arrangements with a construction firm (e.g., Bechtel) about the price the utility will be charged to build the plant. In the long run, the price should cover the entire construction firm's costs, including profits; otherwise, the firm would not be in business. The costs cited above may not include the needed profits. Additionally, the utility itself will incur expenses, often called owner's costs. These would include the cost of the land, licensing/regulatory activities, utility oversight and quality control functions, and startup activities. In total, depending upon the contractual arrangements between the utility and the construction firm, these costs could range from about $50 to $150 per kW. Such expenses may, or may not, be included in the costs reported above.

Third, building a nuclear power plant requires substantial amounts of engineering support activities with costs that could be in excess of $100 million. An example of this would be the design review of a component by a senior staff member that is not directly associated with the project. Some have argued that the reported foreign costs may not include all of these overhead expenses.[3] Additionally, in certain cases in the Far East, residents located in close proximity to a nuclear plant may be directly or indirectly reimbursed to minimize local resistance to its construction, and those expenses could be included in the published costs.[4] If similar activities were not to occur in the United States, the cited costs would remain "too high."

Again, the design of the nuclear power plants whose costs were reported above (i.e., $1700–$2000 per kW) the reflected incremental improvements over plants that became operational in the 1980s to mid-1990s. Some industry analyses suggest that, in the United States, these designs might not be competitive, and as a result they are developing a series of more advanced light-water nuclear power plant designs. These are more simplified and have more passive safety features, and should, therefore, be less expensive to build. That is, operator intervention was needed to activate the safety systems in the 1980s to mid-1990s vintage light-water reactors and because of possible human error/equipment failure, numerous backup safety systems were needed. In the newer designs with more passive safety features, less human intervention is needed and equipment failure is less likely. Thus, fewer backup safety systems are required, reducing construction costs. These advanced designs include the AP1000B, a PWR designed by Westinghouse and another more advanced BWR, designed by GE, called the ESBWR. The former design was just approved by the Nuclear Regulatory Commission (NRC), and the latter is in the initial stages of NRC review.

Unfortunately, none of these units have been built anywhere in the world, so there is no information on realized costs, and, thus, bottom-up cost estimates must be examined. According to the vendors' bottom-up estimates, they think (or hope, rather) that, eventually, plants with these designs could be built for about \$1000–\$1400 per kW.[5] It must be noted that, historically, similar vendor estimates have been too low. Indeed, in the United States, no nuclear power plant has been built on-time and on-budget.[6] Additionally, as can be seen from Fig. 1, these vendor estimates tend to be much lower than the realized costs of any nuclear plant built anywhere in the world. Thus, these cost savings may not be achieved.

In short, until new nuclear plants are built in the United States, construction costs are largely "unknown" and, thus, it is impossible to use any single estimate to accurately predict costs. In the analysis that follows, therefore, a range of costs will be used. This reflects some of the inherent uncertainty with building the first few nuclear power plants in the United States.

ENVIRONMENTAL ISSUES

In the long run, environmental issues are probably the most important set of factors affecting the cost of both nuclear power and its alternatives. Current U.S. federal law requires that all new fossil fuel-fired power plants meet certain standards for sulfur dioxide and nitrogen oxide emissions. The capital cost estimates used in this analysis include the expenditures for the equipment needed to meet these standards. Additionally, in the next section, the cost of complying with the Kyoto Treaty will also be considered, even though the United States has not yet ratified it.

Nuclear power has its own set of environmental problems—namely, the possibility of exposing the public to radiation, decommissioning, and spent fuel disposal. Since the capital cost estimates cited above generally dealt with plants whose designs were certified by the NRC, all of the expenses of meeting NRC safety requirements are included in the costs. Nuclear power plant decommissioning deals with dismantling the plant and restoring the site for other uses once the unit is permanently retired. While plant dismantlement and site restoration are common to all power plants, nuclear

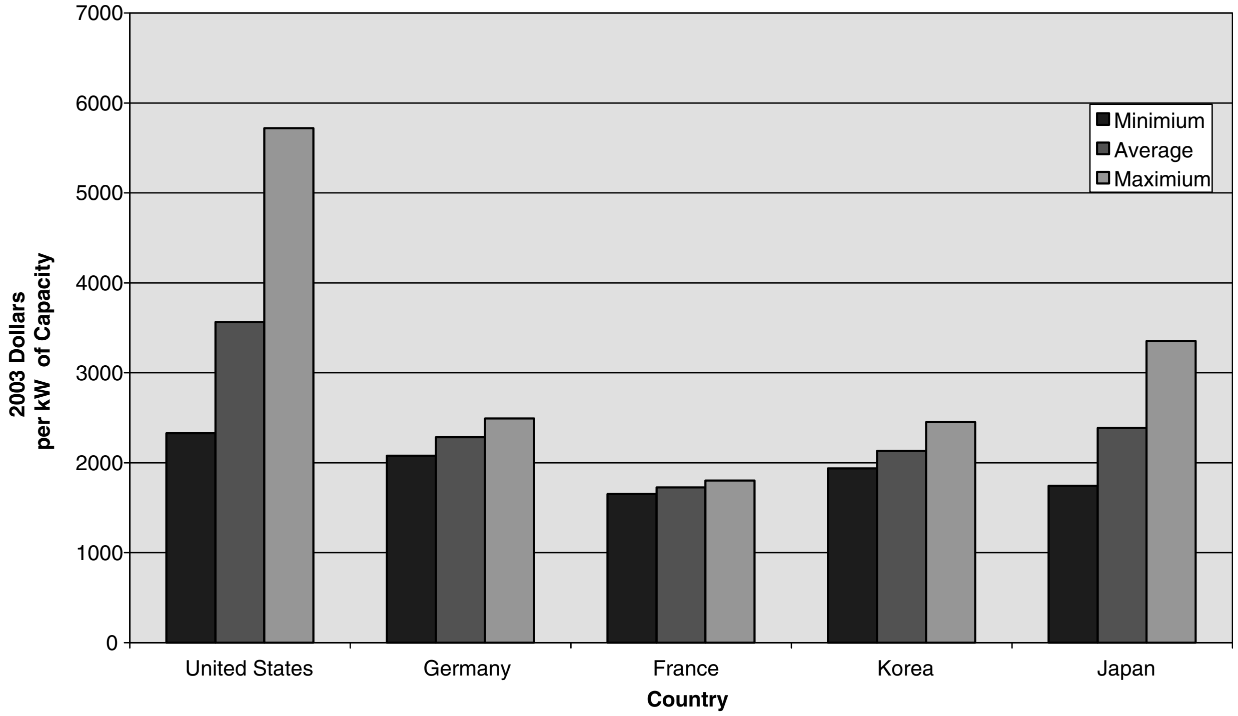


Fig. 1 Realized nuclear construction costs for units entering commercial operation in the 1980s.
Source: From McKerron, Gordon, "Why do nuclear costs continue to increase", *Energy Policy*, July 1992; Chung-Taek Park, "The experience of nuclear power development in the Republic of Korea," *Energy Policy*, July 1992; John Marshall and Peter Navarro, "Cost of nuclear power plant construction: theory and new evidence," *Rand Journal of Economics*, Spring 1991; Energy Information Administration, *An analysis of Nuclear power plant constriction costs*, DOE/EIA-0485, 1986.

power is treated as "different" because many of the components are highly radioactive. Thus, to protect public heath and safety, the NRC also regulates the dismantlement of the plant once it is retired. The NRC must approve a firm's decommissioning plans, insure that the residual site radiation meets NRC requirements, and oversee the actual plant dismantlement. In total, the cost of dismantling a plant and restoring the site is about \$400–\$600 million per unit.

The high-level waste from a nuclear power plant will remain highly radioactive for hundreds of thousands of years, and the permanent disposal of this waste is a major environmental issue affecting the universal acceptance of nuclear power. Over the years, methods ranging from shooting the waste into space to burying it under the sea have been proposed. The United States has chosen another option—namely, to permanently bury the waste underground. Starting in the 1980s, the U.S. Department of Energy (DOE) began to explore a number of sites in the Eastern and Western United States, and is currently concentrating their efforts on one site at Yucca Mountain, Nevada. The site has been deemed to be satisfactory and the DOE is in the process of submitting detailed plans to the NRC for approval. At the earliest, the DOE hopes that Yucca Mountain can begin receiving nuclear waste by 2012, and the site will remain open until about 2110.

Not surprisingly, substantial amounts of opposition by the state of Nevada exist, and there have been a number of court challenges to the proposal by the state. Recently, parts of the regulations affecting the repository's performance have been invalidated by the U.S. Court of Appeals, and the resolution of these issues could delay the opening of the repository to beyond 2012. There is also additional controversy about the repository's design that could result in additional delays.[7,8] Note also that there is a statutory 70,000 metric tons limit to the amount of waste that can be stored at Yucca Mountain. Since this limit is less than the projected amount of waste from existing power plants, at some point Congressional action will be needed to change this constraint.

The DOE has recently estimated that the total cost of transporting the waste to Yucca Mountain and disposing the spent fuel there will be about \$57 billion.[9] Given all the technical uncertainties about the design of the repository and all the political resistance in Nevada, the current cost estimate could possibly be "too low." Additionally, after the spent fuel is removed from the reactor, it is placed into onsite storage pools. In a number of cases, however, the amount of waste generated has exceeded the capacity of these pools. As a result, as of early 2003, waste is being stored in "dry casks" at 19 of the 66 sites with operating power plants. Additionally, even if Yucca Mountain opens on schedule, in the future, additional amounts of waste would have to be stored onsite in dry casks. Dry cask storage costs are substantial and are not included in the \$57 billion estimate. Thus, the total cost could be much greater than \$57 billion.

In total, the estimated decommissioning and high-level waste disposal costs (the back-end costs) are, at a minimum, probably about \$1100 per kW of capacity—an amount roughly equal to the capital cost of a coal plant (the current capacity of all operating nuclear power plants in the United States is about 100 GW. Thus, the total spent fuel disposal costs per kW of capacity would be about \$570 per kW. Since decommissioning costs are about \$500 per kW, the total back-end costs would be about \$1070 per kW (\$570 + \$500)). However, these costs will be incurred many years (decades) into the future, and because of the discounting process, the present value today of these costs is about 1.3–1.4 mills per kilowatt hour (kWh) of plant output (1 mill = 0.1 cents). The DOE is collecting a fee of 1 mill per kWh of electricity generated from nuclear power plants, which is placed in a trust fund that will earn interest. The DOE has estimated that this charge, plus all the accrued interest, will be sufficient to cover the \$57 billion waste disposal costs. Similarly, a charge of about 0.2–0.3 mills per kWh of electricity generated from a 1,000 mW nuclear power plant which is invested in a trust would be sufficient to cover all the \$400–600 million to decommission it. The total costs of generating electricity from nuclear power plants are about 40–65 mills per kWh. Since the back-end costs are incurred over long time periods and because of the "magic of compound interest," the back-end costs today would be about 1%–2% of the total cost.

Because of the discounting process, very large costs imposed on future generations will appear to be very small today. Consequently, some intergenerational equity issues dealing with evaluating the back-end costs have arisen.[10] Some economists have attempted to include intergenerational fairness considerations into discounting formulas.[11] Unfortunately, this research is very controversial and it is virtually impossible to derive quantitative present value estimates using this work. Thus, there are some equity issues dealing with nuclear power that cannot be resolved with economic analysis.

DISCOUNT RATES AND FOSSIL FUEL PRICES

Any power plant will operate for many years, and all the costs incurred in the future must be discounted back to the present. The arithmetic of discounting is trivial—almost all spreadsheets have functions that do the computations automatically. Unfortunately, the same could not be said for the selection of the discount rate, which is one of the most complex (and controversial) subjects in modern financial economics.[12] The changing structure of electricity markets in the United States simply compounds the problems.

Table 1 Other assumptions used in the analysis

Nuclear non-fuel O&M ($s per kW)	55
Nuclear fuel costs (mills per kWh)	4
Nuclear leadtime (years)	5
Capacity factors for all plants (percent)	90
Coal capital costs ($s per kW)	1154
Coal non-fuel O&M ($s per kW)	54
Coal heat rate (Btu/kWh)	9000
Coal leadtime (years)	4
Annual real coal price escalation rate	0
Natural gas capital costs ($s per kW)	608
Natural gas non-fuel O&M ($s per kW)	29
Natural gas heat rate (Btu/kWh)	7000
Natural gas leadtime (years)	3
Annual real natural gas price escalation rate (%)	2
Real after-tax discount rate-high rate (%)	11
Real after-tax discount rate-low rate (%)	4.6

Sources: From Energy Information Administration, *Annual Energy Outlook*, 2004, DOE/EIA-0383(2004), January 2004. Massachusetts Institute of Technology, *The Future of Nuclear Power*, Cambridge, Mass, 2004.

Because of all of these uncertainties, the analysis uses two real discount rates of between about 5 and 11% (refer to Table 1). The 11% real after-tax discount rate assumes that building and operating any power plant is fairly risky, and is roughly consistent with ones used to examine airline industry investments. This rate is also consistent with actual rates used by firms in competitive industries.[13] The discount rate used for completely risk-free investments would be about 3%, which is slightly less than the real 5% after-tax rate. Thus, the 5% rate assumes that power plant investments are relatively safe (from a financial viewpoint) or are being shifted to consumers and/or taxpayers. This rate is also consistent with the ones used in regulated industries.

Another factor affecting the economics of nuclear power, or any industry really, is the cost of the alternative. The basic assumptions about the cost of generating electricity from fossil fuel-fired power plants are also shown in Table 1. Environmental considerations aside, there is little controversy about the fossil fuel capital cost estimates. Additionally, the United States has some of the largest coal reserves in the world and, therefore, there is little uncertainty about future coal prices. Because of the location of the reserves and high transportation costs, there are substantial regional variations in coal prices. Currently, on average, coal costs about $1.30 per million Btu, and ranges from a low of about $.90 to about $1.90 per million Btu.

There is, however, uncertainty about future natural gas prices. As seen in Fig. 2, natural gas prices started to fall shortly after they were deregulated in the early 1980s, and fell to about $2.20 per million Btu in the mid-1990s. However, over the last few years, prices have again increased to about $5.80 per million Btu. Most forecasters think that this increase is temporary, and by about 2010, prices will fall to roughly $4.20 per million Btu; then, they will again increase by about 1%–3% per year.[5]

RESULTS

As was noted above, the costs of building nuclear power plants are highly uncertain. Wide regional variations in coal prices and uncertainties about future natural gas prices also exist. Because of these factors, point estimates of the levelized costs of generating electricity from coal, natural gas, and/or nuclear power plants have little value. The general approach used here is, therefore, to derive various combinations of nuclear capital costs and "current" coal/natural gas prices that would result in nuclear power and coal/natural gas being equally economic (i.e., the levelized costs of coal/natural gas and nuclear are the same) (the method used to compute the levelized costs can be found in Ref. [2]. Current refers to the first year of the plant's operation). Once this is done, a number of "what if" scenarios can be developed. While this is not ideal, it is the best that can be done given all the uncertainties involved.

The results of a comparison of the economics of nuclear power relative to coal using the higher discount rate are shown in Fig. 3. The solid line shows the combinations of nuclear capital expenses and "current" coal prices that would result in both technologies having the same levelized cost. Thus, any combination of nuclear capital costs and coal prices that falls in the region denoted as E would result in nuclear power being economic. Similarly, any combination of these two factors that fall in the region U would result in nuclear power being uneconomic relative to coal. As was noted above, current coal prices range from about $.90 to about $1.90 per million Btu. Thus, if coal prices remain constant in real terms over the next six years, nuclear capital costs would have to be less than or equal to about $1100 to $1500 per kW before nuclear power would be economic (points A and B). The national average coal price is about $1.30 per million Btu. Thus, if the plant is located in a region with average coal prices, nuclear capital costs would have to be less than or equal to $1200 per kW before nuclear power would be economic (point C).

A comparison of the economics of nuclear power relative to gas-fired power plants is shown in Fig. 4. Currently, natural gas prices in the United States are about $6 per million Btu and most forecasters expect that they will fall to about $4.20 per million Btu by the end of the decade. If this occurs, nuclear capital costs would have to be less than or equal to about $1500 per kW before nuclear

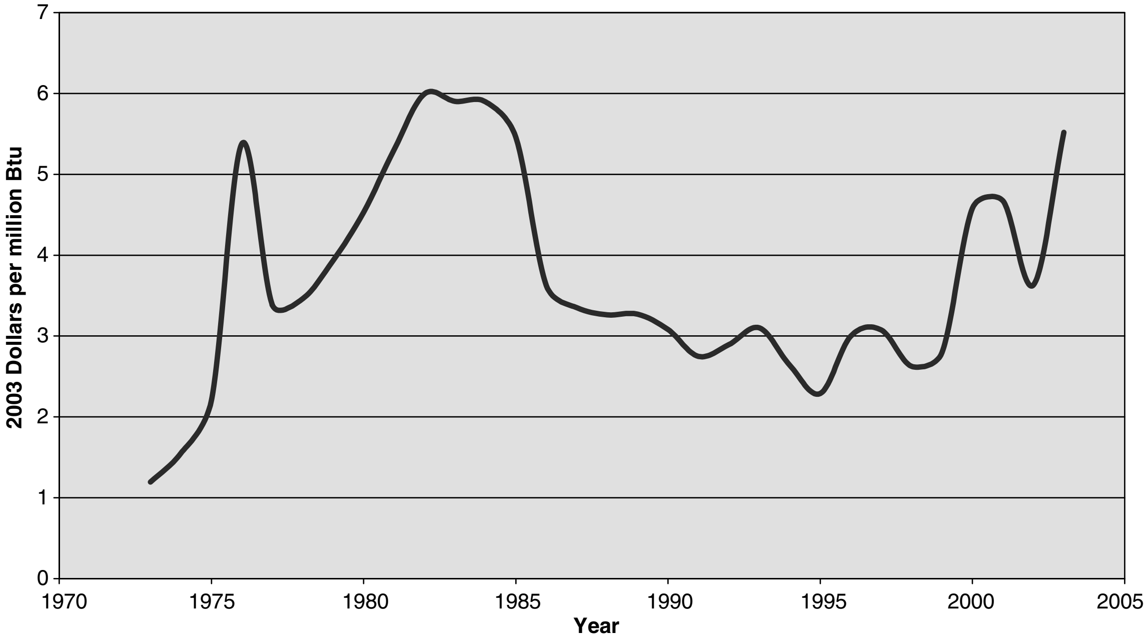


Fig. 2 Cost of natural gas delivered to U.S. electric utilities.
Source: From Energy Information Administration, *Monthly Energy Review*, January 2005.

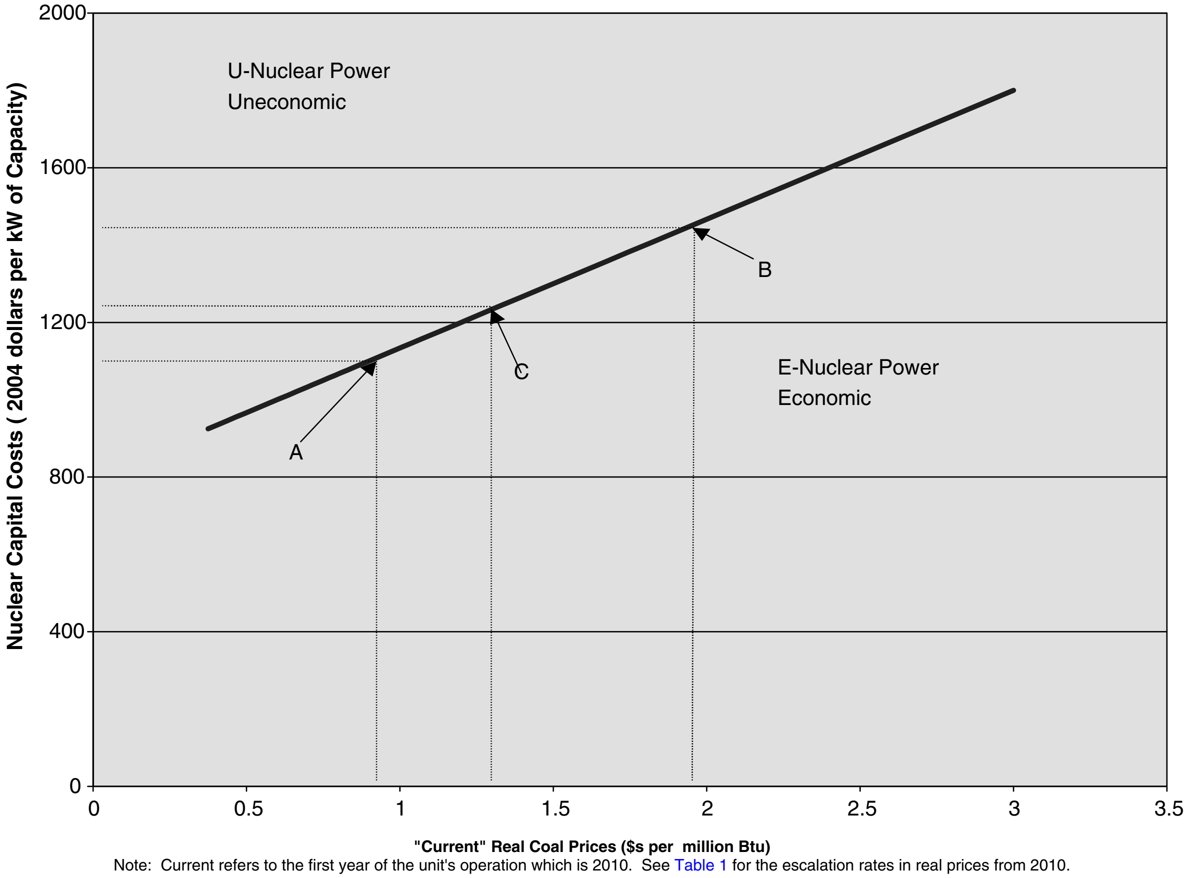


Fig. 3 The economics of nuclear power relative to coal-fired power plants: higher discount rate.

Nat–Perf

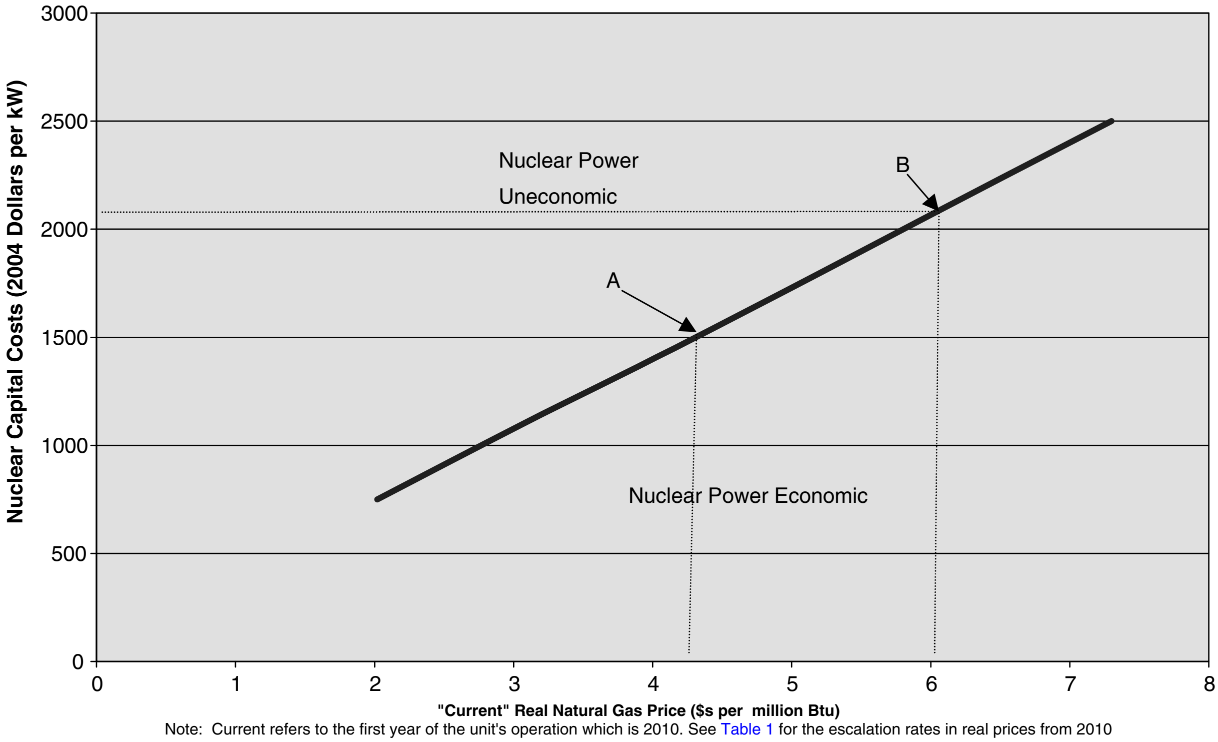


Fig. 4 The economics of nuclear power relative to gas-fired power plants: higher discount rate.

power would be economic relative to gas-fired power plants (point A). If, however, gas prices remain at their current levels for the next six or so years, nuclear power would be economic if the nuclear plant could be built for about $2000 or less (point B).

As discussed above, given all the caveats, most of the available information suggests that the U.S. costs of building nuclear powers plants with designs similar to the ones being constructed in Finland and the Far East are about $1700–$2000. Since this range is much greater than the "break-even" range of nuclear capital costs of about $1100–$1500, nuclear plants with those designs would probably not be economic in the United States. The one exception is if coal is not an alternative because of site-related or environmental factors and gas prices remain at current levels (about $6.00 per million Btu). This general result was, in fact, one motivation for developing the more advanced AP1000 and ESBWR designs. The vendors think, and hope, that the capital costs of the newly designed reactors will be about $1000–$1400. If these expectations are realized, nuclear power would be economic.

These observations assume that building and operating any power plant is relatively risky and that the decision makers will bear all of that risk. If, in fact, this assumption is not correct or if the regulatory/market structure shifts that risk to others, a different set of conclusions could be made. In particular, Figs. 5 and 6 show the economics of nuclear power relative to gas and coal-fired power plants, respectively, using the lower discount rate. Even if gas prices fall, in its current form, nuclear power would be economic relative to natural gas (refer to Fig. 5). If coal is an alternative, it would be economic to build nuclear plants with capital costs of about $2000 per kW or less in regions of the United States that are distant from coal mines and, therefore, have high coal prices (as in point B in Fig. 6). In regions of the United States with average coal prices, nuclear capital costs would have to be about $1400–$1500 per kW or less before nuclear power would be economic relative to coal-fired power plants (refer to point C in Fig. 6). Thus, there is a second case where nuclear power plants with current designs are economic—namely, if the risk of building and operating any power plant is perceived to be low (or to have been shifted to consumers/taxpayers) and the firm is located in areas distant from coal mines.

The fossil fuel costs used in this analysis included all the expenditures needed to meet existing environmental regulations. Since the United States has not ratified the Kyoto Treaty, the costs of reducing carbon emissions have not been considered. A number of studies have estimated the costs of U.S. compliance with the Kyoto agreements. These studies found that a tax on carbon emissions from fossil fuel power plants from about $50 (or less) to greater

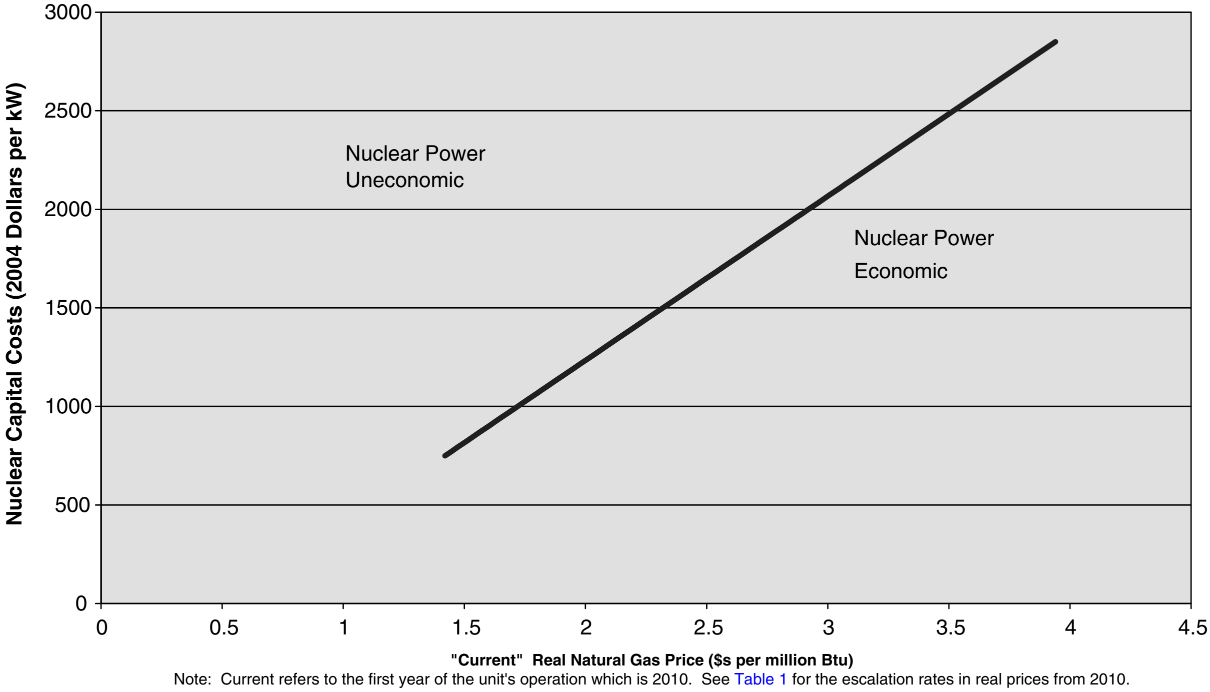


Fig. 5 The economics of nuclear power relative to gas-fired power plants: lower discount.

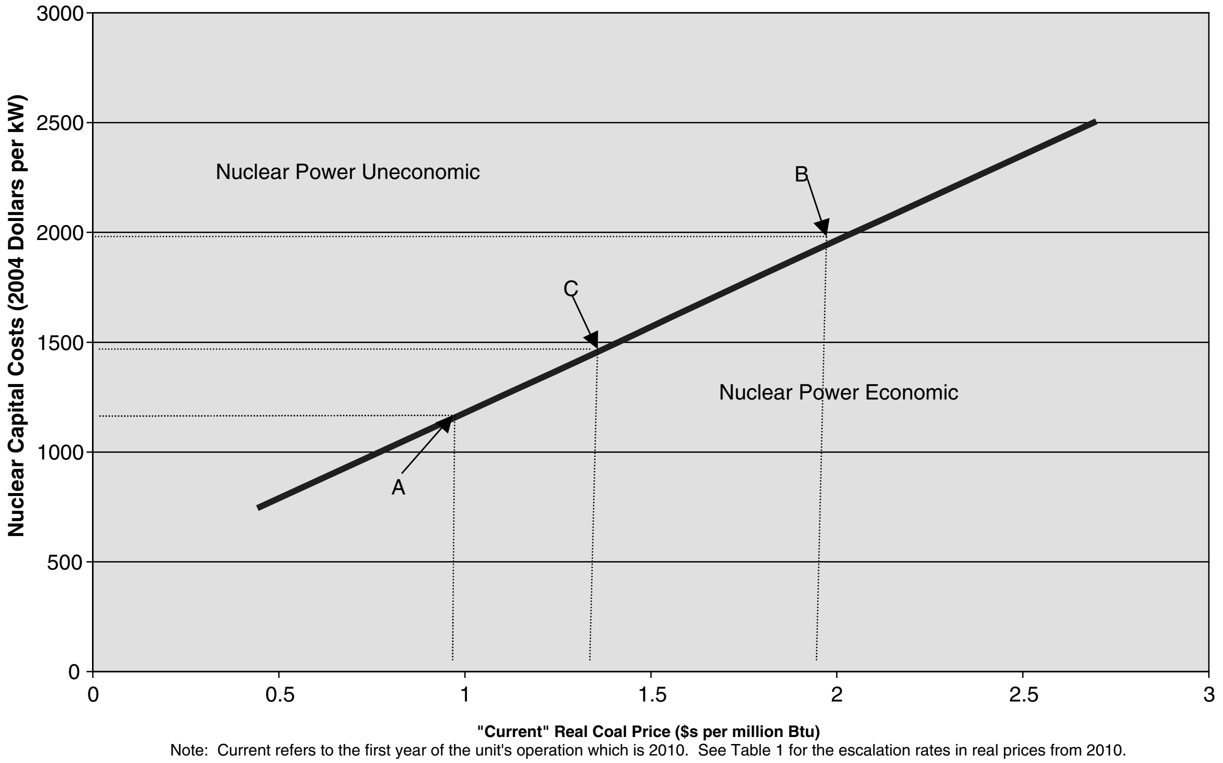


Fig. 6 The economics of nuclear power relative to coal-fired power plants: lower discount rate.

Nat–Perf

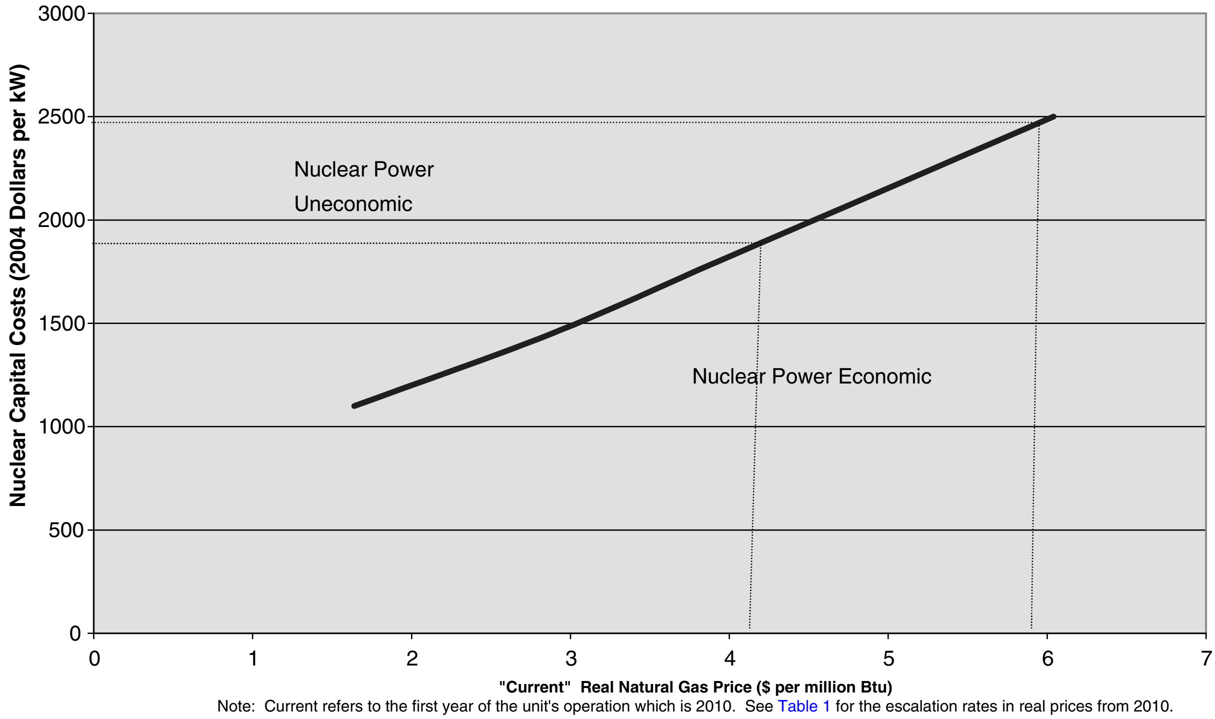


Fig. 7 The economics of nuclear power relative to gas-fired power plants: $100 per ton carbon tax and higher discount rate.

than $150 per ton would be needed.[2] The effects of a $100 per ton carbon tax on the economics of nuclear power will now be considered.

Carbon emissions are much greater from coal than from natural gas-fired power plants, so it was not surprising to find that if a $100 per ton carbon tax was imposed, regardless of the price of coal, nuclear power would be economic. Because these results are so predicatable, for brevity sake, they are not reported. The results of comparisons with natural gas, shown in Fig. 7, suggest that if natural gas prices fall to about $4.20 per million Btu by the end of the decade, nuclear capital costs would have to be less than about $2000 per kW before nuclear power would be economic. This result suggests that if fairly stringent carbon reduction requirements were enacted, current nuclear power plant designs would probably be economic in the United States.

SUMMARY

This article attempted to answer the question: "Is Nuclear Power Economic?" There are two major factors influencing the economics of nuclear power, and since major uncertainties with both of them exist, it is impossible to give an unqualified answer. The most obvious uncertainty is the cost of building a nuclear power plant. Environmental considerations aside, analysis has found that, generally, nuclear capital costs must be less than $1500 per kW before nuclear power would be economic. If the costs of recently completed plants in Asia and Finland are at all indicative of the cost of building plants with similar designs in United States, it probably would not be economic to do so. Indeed, the nuclear industry also recognizes this point, and, therefore, they are designing more advanced light-water reactors that are simpler and include more passive safety features. The vendors think (or hope, rather) that nuclear power plants with these new designs could be built for about $1000–$1400 per kW. If, in fact, cost reductions of this order of magnitude are achieved, nuclear power would be economic.

The second factor is the cost of the alternative which would be either coal-fired power plants or combined-cycle, natural gas units. The major uncertainty here is the enactment of policies that limit carbon emissions. The economics of nuclear power would be greatly improved if all of the external costs related to global warming were included in the cost of generating electricity from fossil fuel-fired power plants. It must be noted that nuclear power has its own set of environmental costs in the form of nuclear waste disposal and decommissioning. In total, these expenses are probably about the same as the capital costs of a coal plant, but since they will be incurred over hundreds of years, and because of discounting, the back-end costs are very small "today." Some have objected to discounting expenses incurred over very long time periods because this procedure represents a strong incentive to impose large

costs on future generations. However, such equity considerations are outside the realm of economic analysis.

ACKNOWLEDGMENTS

I would like to thank my past and present colleagues at the Energy Information Administration for their many useful discussions with me about the economics of nuclear power and their assistance in my research in studying this subject. However, the view and opinions stated in this article are the author's alone and do not represent the official position of the Energy Information Administration or the United States Department of Energy. All errors are the sole responsibility of the author.

REFERENCES

1. Merrow, E.W.; Phillips, K.E.; Myers, C.W. *Understanding Cost Growth and Performance Shortfalls in Pioneer Process Plants*; The Rand Corp: Santa Monica, CA, 1981.
2. Deutch, J.; Moniz, E. *The Future of Nuclear Power*. Massachusetts Institute of Technology: Cambridge, Mass, 2004.
3. Thomas, S. *The Realities of Nuclear Power; International Economic and Regulatory Experience*; Cambridge University Press: New York, 1988.
4. Lowinger, T.; Hinman, G. *Nuclear Power at the Crossroads*; International Research Center for Energy and Economic Development: Boulder, CO, 1994.
5. *Annual Energy Outlook, 2004*. DOE/EIA-0383(2004); Energy Information Administration: Washington, DC, 2004.
6. *An Analysis of Nuclear Power Plant Construction Costs*. DOE/EIA-0485; Energy Information Administration: Washington, DC, 1986.
7. *Report to the U.S. Congress and Secretary of Energy*. US Nuclear Waste Technical Review Board (NWTRB): Arlington, VA, 2004.
8. Craig, P.R. Rush to judgment at Yucca mountain. Sci. Democratic Action IEER **2004**, *12* (3), 1–10.
9. *Analysis of the Total System Life Cycle Cost of the Civilian Radioactive Waste Management Program*. DOE/RW-0533; Office of Civilian Radioactive Waste Management (OCRWM), U.S. Department of Energy: Washington, DC, June 2001.
10. Surry, J. Ethics of nuclear decommissioning. Energy Policy **1992**, *20* (7), 632–640.
11. Lind, R. *Discounting for Time and Risk in Energy Policy*; John Hopkins University Press: Baltimore, MD, 1982.
12. Shapiro, A. *Modern Corporate Finance*; MacMillan Publishing Company: New York, 1990.
13. Poterba, J.; Summers, L. A CEO survey of U.S. companies' time horizons and hurdle rates. Sloan Manage. Rev. **1995**, 43–53.

Nat–Perf

Nuclear Energy: Fuel Cycles

James S. Tulenko
Laboratory for Development of Advanced Nuclear Fuels and Materials, University of Florida, Gainesville, Florida, U.S.A.

Abstract
The nuclear fuel cycle utilizes either uranium or thorium, which are relatively plentiful materials. Both materials must undergo several processing steps in order to be converted into a useful fuel for nuclear reactors. For uranium, this involves conversion to UF_6, enrichment, and final processing into fuel elements. Thorium is more complex and must first be irradiated in a reactor before a useful fuel is formed.

INTRODUCTION

The nuclear fuel cycle is powered by the release of 200 million electron volts of energy every time a single uranium atom is fissioned. The enormity of this energy release can be better understood when one learns that the energy release from the combustion of a carbon atom to carbon dioxide is four electron volts. Thus, the energy release per atom in the nuclear fuel cycle is 50 million times more than in the fossil (carbon) fuel cycle. It is this 50 million times difference that makes the nuclear fuel cycle relatively inexpensive once the nuclear power plant is built.

For nuclear power to be a sustainable power source, it is important that nuclear fuel, which powers the nuclear fuel cycle, is readily available and sustainable. The nuclear fuel cycle uses two naturally occurring elements that are both relatively common metals and obtained through mining: uranium and thorium.

Uranium is a common constituent of the earth's crust and it occurs in most rocks in concentrations of two (sedimentary rocks) to four (granite) parts per million. Uranium also occurs in seawater in a concentration of 0.003 parts per million, which corresponds to approximately four billion tons of uranium in the oceans. Uranium (1.8 g/tn) is more abundant than common materials such as silver (0.07 g/tn), tungsten (1.5 g/tn), and Molybdenum (1.5 g/tn).[1] It is 800 times more abundant than gold. Canada is the largest producer of uranium, with Australia being a close second. In terms of known uranium resources, Australia has the most followed closely by Canada and Kazakhstan. Natural (as in mined) uranium contains in atomic abundance: 99.2745% Uranium-238 (U238); 0.72% Uranium-235 (U235); and 0.0055% Uranium-234 (U234). Uranium has atomic number 92, meaning all uranium atoms contain 92 protons and electrons, with the rest of the mass number being composed of neutrons. All uranium isotopes are radioactive. This radioactive property makes the detection of uranium deposits relatively easy, even allowing for prospecting by air. Uranium-238 has a half-life of 4.5×10^9 years (4.5 billion years), U-235 has a half- life of 7.1×10^8 years (710 million years), and U-234 has a half-life of 2.5×10^5 years (250 thousand years). All of the U234 currently present comes from the decay chain of U-238. Uranium-235 is the only fissile (fissile means materials can be fissioned by thermal neutrons) isotope available in nature. Uranium can be a fissionable fuel as mined for use in the CANDU reactors developed in Canada. In these reactors, the water moderator is enriched to contain only pure deuterium. As a by-product of the operation of a nuclear reactor, U238 absorbs a neutron to form the fissile fuel, plutonium-239. Other fissile isotopes are uranium-233 (from thorium 232) and plutonium-241. Note that it is the odd number isotopes—233, 235, 239, and 241—that are fissionable.

Thorium is even more readily available for nuclear fuel, being four times more abundant than uranium in the earth's crust. Thorium is the 39th most common element in the earth's crust and it is about as common as lead. Thorium is present in the earth's crust with an average concentration of about 9.6 parts per million. Thorium must be converted in a nuclear reactor by absorption of a neutron into the fissionable fuel, uranium-233. Thorium has only one naturally occurring isotope, thorium-232, which is radioactive with a half-life of 1.3×10^{10} years. India, which has large thorium deposits, has been a leader in utilizing thorium to breed uranium-233 to serve as a nuclear fuel. Other countries with major deposits of thorium are Australia, Norway, and the United States.

A NATURAL REACTOR

The formation of the planets and the sun from an original cloud of dust and gas may have been triggered by the

Keywords: Uranium; Thorium; Uranium mining; Uranium milling; Uranium conversion; Uranium-235 enrichment; Nuclear fuel fabrication.

explosion of a nearby supernova. Only a supernova can manufacture elements heavier than iron, including uranium. With a half-life of 710 million years, U-235 started out making up nearly half of all uranium when the solar system began some 4560 million years ago. Over time, the percentage of U235 in natural uranium has steadily decreased. However, there is evidence that a natural nuclear reactor was formed approximately 1.8 billion years ago in Oklo in Ghana, Africa: a river flowed above a buried rich uranium ore body and the water acted as a moderator to the enriched uranium ore body, which had a natural enrichment of about 3.2% at that time. The nuclear interactions reached the critical point and the reactor went critical (produced sustainable power). As the reactor heated up, the water turned to steam and boiled away. With the moderator gone, the chain reaction stopped and did not start again until the ore body cooled and the water returned. This simple feedback cycle kept the Oklo reactors (there were at least a dozen) active until the U-235 was depleted. The reactors operated for about a million years. When the Oklo mine was producing ore in the 1970s, it was that telltale depletion of the U-235 concentration, unheard of in nature, which tipped scientists off as to its origin. The finding of the fission products confirmed the operation of the natural nuclear reactors.

URANIUM HISTORY

In 1789, Martin Heinrich Klaproth discovered uranium in the mineral pitchblend, which is primarily a mix of uranium oxides. No one could identify this newly isolated material, so in honor of the planet Uranus, which had just been discovered, his new material was called uranium. Although Klaproth (as well as the rest of the scientific community) believed that the substance he extracted from pitchblend was pure uranium, it was actually uranium dioxide (UO_2). It wasn't until 1842 that the French chemist Eugene-Melchoir Peligot noticed that pure uranium reacted oddly with uranium tetrachloride (UCl_4). He then proceeded to isolate pure uranium by heating the UO_2 with potassium in a platinum crucible. Radioactivity was first discovered in 1896 when the French scientist Henri Becquerel accidentally placed some uranium salts near some paper-wrapped photographic plates and discovered the natural radioactivity of uranium by noticing the exposure of the covered photographic plates.

Uranium compounds have been used for centuries to color glass. Uranium trioxide (UO_3) was used in the manufacture of a distinctive orange Fiesta® dinnerware. In 1938, Otto Hahn (1879–1968), Lise Meitner (1878–1968), and Fritz Strassmann (1902–1980) recognized that the uranium atom, under bombardment by neutrons, actually split or fissioned.

When a uranium or plutonium atom is fissioned, it releases 200 million electron volts of energy, while the burning of a carbon (coal) atom releases four electron volts. This difference of 50 million times the energy-release shows the tremendous difference in magnitude between chemical and nuclear energy.

Thorium was discovered in 1829 by the Swedish chemist Jons Jacob Berzelius, who named the element after Thor, the mythical Scandinavian god of war. Berzelius also was the first to isolate cerium, selenium, silicon, and zirconium. Thorium and thorium compounds have very high melting temperatures. As a result, thorium was used for high-temperature applications such as coatings on tungsten filaments in light bulbs and for high-temperature laboratory equipment. However, its use outside the nuclear fuel cycle has been greatly diminished because of state and federal laws concerning the handling and disposal of radioactive materials. Thorium is found in the minerals monazite and thorianite.

MINING AND MILLING

The earliest recovery of uranium was from pitchblend, an ore with a very high UO_2 content of up to 70%. Pitchblend also contains radium, thorium, cerium, and lead. It is mostly found with deposits that contain phosphates, arsenates, and vanadates. Uranium exists in nature in two valence states, U^{6+} and U^{4+}. These properties are key proponents in the geological distribution of uranium. U^{6+} is soluble in water, but changes to the insoluble U^{4+} in a reducing environment. The occurrence of reducing environments in riverbeds and seas have led to the formation of rich uranium deposits. A rich uranium deposit contains 2% uranium and economic deposits are as low as 0.1%. Once the ore is mined, it is sent to a mill, which is really a chemical plant that extracts uranium from the ore. The ore arrives via truck and is crushed and leached, and approximately 90%–95% of the uranium is recovered through solvent extraction. During processing, a large waste stream called tails is formed, containing approximately 98%–99.9% of the ore mined. Because this waste stream or tails contains all the radioactive daughter products of uranium, such as radon and radium, this waste stream must be carefully controlled and stabilized. The tailings pile must have a cover designed to control radiological hazards for a minimum of 200 years and up to 1000 years to the greatest, reasonably achievable extent. It must also limit radon (^{222}Rn) releases to 20 pCi/m^2/s averaged over the disposal area. The end uranium product of the milling process is U_3O_8, better know as "yellowcake" because of its color.

URANIUM CONVERSION AND ENRICHING

The U_3O_8 concentrate must be both purified and converted to uranium hexafluoride (UF_6), which is the form required for the enriching process. At the conversion facility, the uranium oxide is combined with anhydrous HF and fluorine gas in a series of chemical reactions to form the chemical compound UF_6. The product UF_6 is placed into steel cylinders and shipped as a solid to a gaseous diffusion or gaseous centrifuge plant for enrichment. UF_6 is a white crystalline solid at room temperature (its triple point is 64°C [147.3°F]) and it sublimes (changes state directly from a solid to a gas) at 56.5°C (133.8°F) at atmospheric pressure. The liquid phase only exists under pressures greater than about 1.5 atmospheres and at temperatures above 64°C. In the enrichment plant, the solid uranium hexafluoride (UF_6) from the conversion process is heated in its container until it becomes a liquid. The container becomes pressurized as the solid melts and UF_6 gas fills the top of the container. The gaseous diffusion process is based on the difference in rates at which the fluorides of U235 and U238 diffuse though barriers. The uranium that has penetrated the barrier side is slightly enriched in U-235 and is withdrawn and fed into the next higher stage, while the slightly depleted material inside the barrier is recycled back into the next lower stage. It takes many hundreds of stages, one after the other, before the UF_6 gas contains enough U235 to be used in a reactor. Each barrier has millions of holes per square inch, with each hole approximately 10^{-7} in. in diameter. This process is very energy-intensive, as the gas is compressed and expanded at each stage.

The other commercial enriching process, which uses an order of magnitude less energy, is the gaseous centrifuge process. The gas centrifuge uranium enrichment process uses a large number of rotating cylinders in series and parallel formations. Centrifuge machines are interconnected to form trains and cascades. In this process, UF_6 gas is placed in a cylinder and rotated at a high speed. This rotation creates a strong centrifugal force, so that the heavier gas molecules (containing U-238) move toward the outside of the cylinder and the lighter gas molecules (containing U-235) collect closer to the center. The stream that is slightly enriched in U-235 is withdrawn and fed into the next higher stage, while the slightly depleted stream is recycled back into the next lower stage. At each stage of the gaseous diffusion process, the U-235 is enriched by a factor of 1.004, while at each stage of the gaseous centrifuge process, the stage enrichment factor is approximately 1.1–1.2 for a 1-ft diameter rotor spinning at 350 m/s at a temperature of 300 K. For 1 kg of uranium enriched to 5% U-235, 9.4 kg of natural uranium feed are required and 8.4 kg of depleted uranium (tails) with a U-235 isotope content of 0.2% are produced as a waste stream. The U.S. Nuclear Regulatory Commission has deemed depleted uraniuma low-level waste. The Department of Energy (DOE) has over 560,000-metric-tonne stockpiles of uranium tails stored as UF_6 in steel cylinders. The tails uranium has minor uses as a shield for radioactive sources, such as the penetrator in armor piercing shells, a yacht hold ballast, or as a weight for the balancing of helicopter rotor tips and passenger aircraft.

NUCLEAR FUEL FABRICATION

The enriched UF_6 is transported to a fuel fabrication plant, where the UF_6, in solid form in containers is again heated to gaseous form, and the UF_6 gas is chemically processed to form UO_2 powder. This powder is then pressed into pellets, sintered into ceramic form, loaded into Zircaloy tubes, pressurized with helium, and sealed. The rods are then placed into an array bound together with guide tubes, spacer grids, and top and bottom end fittings, which forms the fuel assembly. Depending on the type of light water reactor, a fuel assembly may contain a 9×9 or 17×17 array of fuel rods and have dimensions of 5–9 in. square by about 12 ft long. The fuel is placed into containers and trucked to the nuclear fuel plants to generate electricity. A single pressurized water fuel assembly contains about 500 kg of enriched uranium and can produce 200,000,000 kW h of electricity. Because the average national electrical yearly use per person is about 12,000 kW h, a single nuclear fuel assembly provides for over 5500 people during their yearly electric needs over its three years of operation.

NUCLEAR FUEL OPERATION AND DISPOSAL

Every 12–24 months, U.S. plants are shut down and the oldest fuel assemblies are removed and replaced with new fuel assemblies. At the end of its useful life, the spent fuel assembly is placed in a borated water storage pond for removal of the radioactive decay heat. After approximately five years of wet storage, the fuel assembly can be removed to dry storage in concrete or steel containers. Because only approximately 5% of the uranium fuel is destroyed, in Europe and Asia, the spent fuel is reprocessed and the 95% of uranium remaining is recycled, with the 5% of radioactive waste products sent to waste storage. At the current time, the United States' policy is to store the spent fuel elements in a waste repository being built at Yucca Mountain in Nevada. The DOE is required by law to be responsible for the spent fuel and collects a fee of 1 mi (one thousandth of a dollar) per kilowatt-hour of nuclear electricity delivered, which is paid by consumers of nuclear-generated electricity. The one assembly described above would generate approximately $200,000 in the waste fund for its disposal.

There is enough uranium and thorium in the world to breed the amount of fuel required to allow nuclear plants to produce the current rate of electrical energy usage for the next 1000 years.

When safely run, the nuclear fuel cycle presents a dependable, non-polluting power sources with a very small waste stream. Currently, 20% of the electrical power needs of the United States and 18% of the world's electrical power needs are met with nuclear energy. Certain events, such as the Chernobyl accident in the Ukraine in 1986, have created awareness of the attention to safety that must be constantly maintained at nuclear plants in order to assure that that the nuclear fuel cycle continues to safely serve the electrical power needs of the world.

REFERENCES

1. Cochran, R.C.; Nicholas, T. *The Nuclear Fuel Cycle: Analysis and Management*; American Nuclear Society: La Grange Park, Illinois, 1999.

BIBLIOGRAPHY

1. U.S. Geological Survey, Mineral Commodity Summaries, January 1999.
2. Bodansky, D. *Nuclear Energy: Principles, Practices and Prospects*; Springer: New York, 2004.
3. http://en.wikipedia.org/wiki/Nuclear_power (accessed March 2007).
4. http://www.world-nuclear.org/education/nfc.htm (accessed March 2007).
5. http://www.uic.com.au/uran.htm (accessed March 2007).

Nuclear Energy: Power Plants

Michael Burke Wood
Ministry of Energy, Al Dasmah, Kuwait

Abstract
A brief history of the main nuclear power plant types, including the reasoning behind fuel, moderator, and coolant choices and their relative advantages and disadvantages, are given in this entry. The three most successful—the pressurized water reactor (PWR), the boiling water reactor (BWR), and the Canadian deuterium uranium (CANDU), as well as a fast reactor—are described in greater detail. Medium-term developments of thermal fission reactors and the long-term possibility of fusion reactors are also described. Nuclear technology and the nuclear fuel cycle are discussed in companion entries.

INTRODUCTION

After rapid growth until the 1980s, a combination of factors has effectively prevented new nuclear power plant construction in the West for two decades. These include the slowing economic demand of the 1980s and the effect of accidents and lack of progress on waste disposal on public confidence, along with power industry uncertainty due to electricity supply deregulation, high capital cost, and increased competition from natural gas. This is not the case in former Soviet countries and the Far East, where construction continues. Nuclear power continues to generate about 16% of the world's electricity. There is growing recognition that irrespective of efficiency in generation, transmission, and end use, nuclear power is the only means of extending developed world standards of per-capita energy consumption to those of the developing world while controlling CO_2 emissions. This is evidenced by the acceleration of nuclear power development in these countries. Fast breeder reactors have not become predominant as expected because of unexpected and greater uranium availability, but they retain a longer-term importance related to waste disposal. Development is improving the safety of established reactor types, but also renewing interest in inherently safer designs and producing simpler, safer, and smaller reactors suitable for local electricity distribution grids in the developing world. The long-term prospect of unlimited cheap energy with negligible environmental consequences has been advanced by the international thermonuclear experimental reactor (ITER) fusion power project.

Keywords: PWR; BWR; CANDU; Gas-cooled reactor; Fast reactor; Advanced reactors; Fusion power.

LIGHT WATER REACTORS

As part of the Manhattan nuclear weapons project, large plutonium production reactors were developed at Oak Ridge National laboratory and production versions built at Hanford in 1945. These were graphite-moderated, horizontal fuel channel designs. The initially-intended helium gas cooling proved inadequate and once-through water cooling was used. However, U.S. civil power reactors had a different military origin: submarine propulsion. A nuclear reactor offered independence from air supply, very infrequent refueling, and, therefore, continuous submerged operation. The application required the smallest, lightest possible reactor. Highly enriched fuel to increase power density and ordinary water as moderator and coolant were chosen. Boiling was suppressed by operation above saturation pressure because of the potential for unstable moderation and fuel overheating if film boiling developed. Thus, a heavy steel reactor vessel and a secondary loop at lower pressure to supply the steam turbine were required. This design was used for USS Nautilus, the first nuclear submarine, in 1953. In 1957, it was used for the 60 MW nuclear power station at a Shippingport near Philadelphia, the first in the United States. It remains the most common and commercially successful today, i.e., the pressurized water reactor (PWR) with about 270, generating 250 GW(e), mainly in the United States, France, Japan, Russia, and Germany. After unsuccessful gas-cooled reactors, France developed PWRs that now produce about 75% of the country's electricity. The Soviet Union turned from the infamous reaktor bolshoy moshchnosti kanalniy (RBMK) to its own PWR, the vodo-vodyannoy energetcheskiy reactor (VVER), which is now the most common design in the former Soviet block. Experiments in the early 1950s proved a reactor with stable nucleate boiling in the core, allowing the boiling water reactor (BWR) to be developed. An experimental version was the first nuclear power plant to supply electricity to a town in 1955. Despite its apparent simplicity, because of the secondary systems required, the BWR competes closely with the PWR on cost and is the second-most common design, with about 90 reactors of 85 GW(e), mainly in the United States, Japan, Germany, and Sweden. PWR and BWR are known as light water reactors to distinguish them from those using heavy water containing deuterium. Both have inherently safe shutdown characteristics because of negative void and power coefficients. Emergency core cooling for decay heat removal in a loss of coolant accident (LOCA) has been more controversial, but concern is now allayed by the evident performance of the safety system in accidents, better computer simulations, and probabilistic fault tree analysis, though the imperative for better safety systems continues.

PRESSURIZED WATER REACTOR

Because it generates steam at only 295°C, the efficiency of the PWR is about 32%, compared with over 40% for some gas-cooled reactors. It can now achieve burn-up rates of up to 40 GW-days/ton. The following description is typical (see Fig. 1). The core is an open matrix of 120 vertical fuel assemblies surrounded by a cylindrical core barrel. Each assembly is about 4 M high each with a matrix of 17 by 17 fuel rods. These are UO_2 pellets 10 mm in diameter contained, clad, in seal-welded tubes of a pure zirconium alloy, zircoloy. Volume for buildup of fission products is provided. To maintain element geometry, the rods are clipped to grids, which also mix the coolant. About 25% of the assemblies have an intermeshing matrix of 4 by 4 control rods of a silver, indium, and cadmium alloy clad in stainless steel, which slide through guides welded to grids. Rod drives are through the pressure vessel head and include magnetic latches, which allow the rods to fall into the core in emergency. Primary coolant enters the reactor vessel at about 275°C and flows downward between the vessel and the core barrel then upward through the core, exiting to the steam generators at about 320°C. There may be two to four primary loops and steam generators, depending on reactor power. The most common steam generator design is shown. Primary coolant from the reactor enters one side of the bottom head then travels through inconel heat exchange U tubes to the other side of the head, then, via main coolant pumps, back to the reactor. From the turbine, steam is condensed and passes though the feed train and feed pump to the steam generator at about 220°C and 70 Bar. It enters above the tube bundle, mixes with water recirculated from the steam separators in the upper head, then circulates through the down comer channel between the shell and the tube bundle wrapper, then upward over the tube bundle, where it boils. It is then dried by the separators and supplied to the turbine dry at 295°C. Circulation in the reactor vessel is driven by the density difference between the steam-water mixture around the tube bundle and the water in the

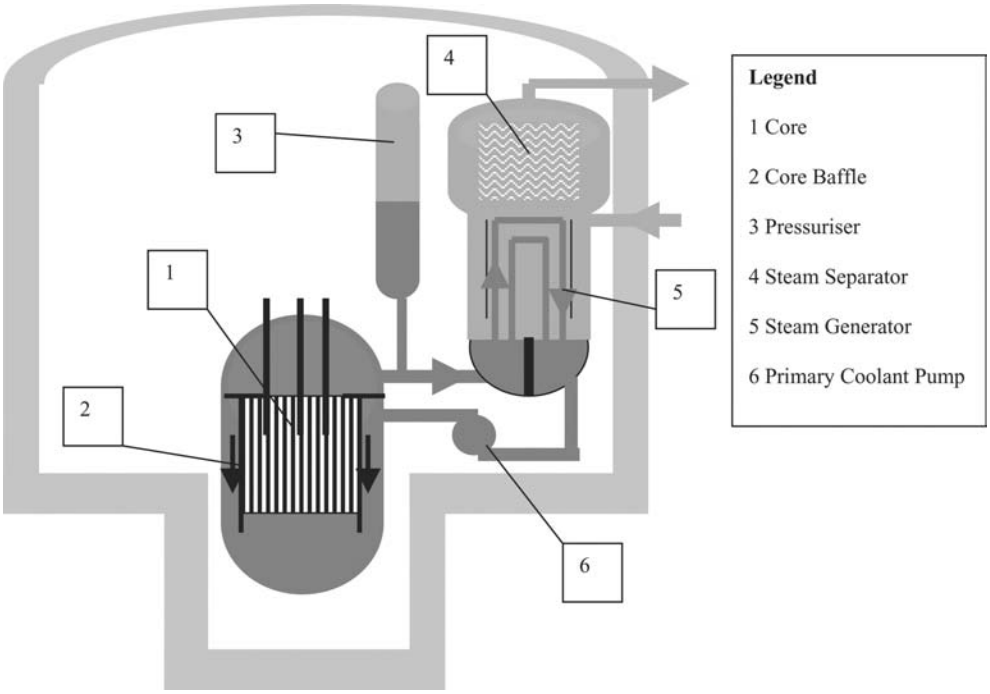


Fig. 1 Pressurised water reactor.

Nat–Perf

down comer. An alternative once-through steam generator, which produces a low degree of superheat, eliminating steam separators, is used in some PWRs. The pressurizer is connected to one hot leg of the primary circuit and maintained at 345°C to ensure a trapped steam bubble, which accommodates primary circuit volume changes caused by temperature variations or liquid moderator injections. It controls pressure by alternatively increasing the steam bubble using electric heaters or reducing it by spraying water drawn from a cold leg of the primary circuit. A relief valve and a relief tank accommodate pressure surges, which are beyond the capacity of the pressurizer. Start-up, rapid transients, and emergencies are controlled by rod operation. Slow reactivity control required by changes in moderator temperature, xenon poisoning, and the build-up of fission products is by chemical shim, the injection of a soluble neutron absorber (normally boric acid) into the primary coolant. When uniformly distributed through the core, shim provides more precise control of fuel burnup than is possible with discrete rods, thus improving operating economics. It also enables the cost and complexity of the control rod system to be reduced. The charging and letdown system controls shim concentration by injecting concentrated boric acid solution or pure water, whichever is required, and compensates volume by removing coolant. The reactor pressure vessel is a massive fabrication of forged rings, flanges, and heads up to 280 mm thick and designed to contain 155 bars. It is impossible to design penetrations through its heads to allow fuel assemblies to pass through and simultaneously seal pressure. Refueling must therefore be off load. It is a major engineering undertaking that involves flooding the concrete well surrounding the reactor vessel to provide cooling and radiation shielding and dismantling a massive bolted flanged joint to permit removal of the upper head. The process is carried out at 12–18 month intervals and about one-third of the fuel is changed in a concentric pattern. Because of these difficulties, submarine PWRs are designed for refueling at more than 10-year intervals, or, in some recent designs, no refueling in the design lifetime. Excess fuel is loaded and compensated by a "burnable poison," or neutron absorber (normally gadolinium), which decays in service as the fuel is burned. This is also used in civilian PWRs. The PWR has negative temperature and void coefficients and therefore shuts down passively without control action in the event of a LOCA. However, decay heat comprising about 8% of reactor power and over 300 MW(th) in a typical large reactor, continues to be produced. Heat transfer for its removal is greatly reduced in a LOCA because primary coolant flashes to steam as pressure is lost. This would result in very rapid failure of the fuel cladding and fission product release. Three diverse emergency core cooling systems (ECCS) are therefore provided. The passive system comprises two or more large, pressurized accumulator tanks above the reactor and connected to the primary circuit at several locations. They are filled with cold borated water, which is automatically released into the reactor if pressure drops. Either of the two further active systems provides replacement cooling water, depending on whether a large breach, which results in major depressurization, or a smaller leak occurs. A containment sump retains cooling water for reuse. The effectiveness of these systems was the subject of a major controversy in the 1970s due to a lack of representative testing and the difficulty of computer simulations. Concerns have been allayed by the evident effectiveness of systems in the Three Mile Island LOCA and by improved analysis methods. Steam is admitted in a dry, saturated form to the steam turbine. Design for nuclear applications is described in the entry on steam turbines.

BOILING WATER REACTOR

The BWR (Fig. 2) is the second-most common design. Its safety characteristics, steam cycle conditions, and efficiency are similar to those of the PWR. Because boiling is permitted in the core, pressure is half that of the PWR, or about 70 bars. A larger vessel is required; therefore, the thickness is 170 mm. Fuel rods are similar to those of the PWR, assembled in smaller matrices, and canned in a zirconium channel, which directs coolant over the rods. This is necessary because of the control rod configuration. These are cruciform assemblies of stainless tubes containing boron carbide, which are inserted into the space between four fuel channels. Because of the steam separators at the top, they are controlled through the bottom head, which is also the optimum location because steaming at the top of the core reduces moderation and displaces neutron flux downward. Chemical shim cannot be used because of the effect of boric acid on steam turbine integrity; therefore, control is by recirculation, as described below, and control rods. Light water coolant is pumped into the reactor vessel by the feed pump, mixes with saturated water recirculated from the steam separators, is force-circulated through the down comer annulus surrounding the core to the bottom of the vessel, and boiled as it passes back up through the core. Force circulation is by jet pumps that are energized by an external pump loop. Steam quality, the ratio of vapor to total mass, leaving the core is kept low to avoid the possibility of film boiling, which reduces moderation, core heat transfer, and causes fuel overheating. Circulation ratio, the ratio of the liquid circulating to the steam produced, is high. Because of this, fuel spacing is greater than in a PWR and a larger pressure vessel diameter is required. From the core, the wet steam enters separators and driers in the vessel head and slightly wet steam is supplied to the

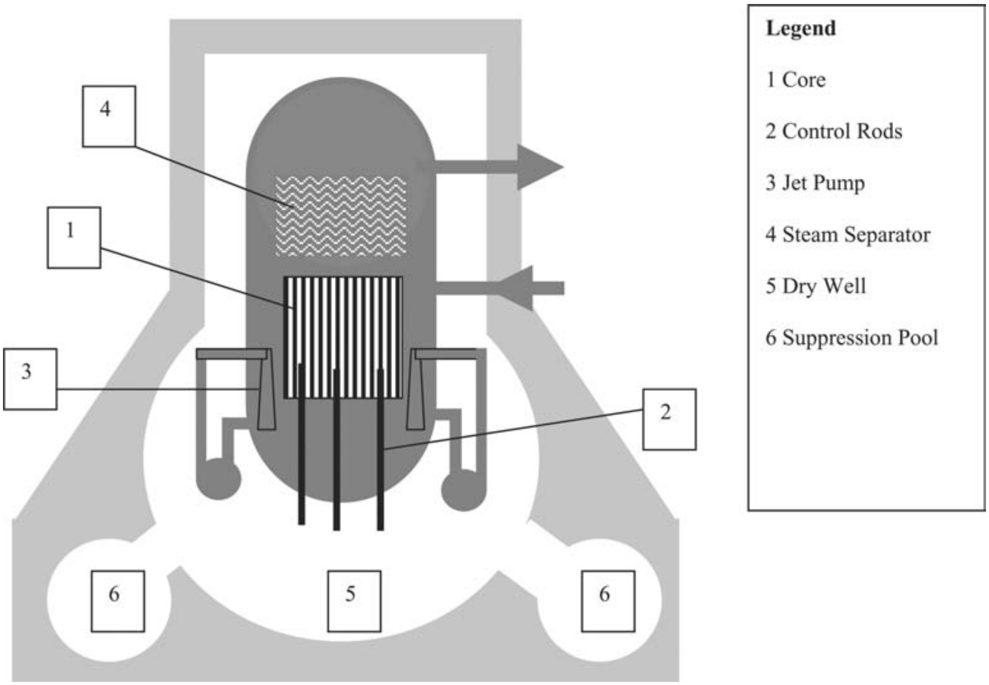


Fig. 2 Boiling water reactor (BWR).

turbine at about 285°C. Because of the presence of steam in the core, a pressurizer is not required. The use of reactor coolant as working fluid has several implications. When turbine load is increased, control valves open to provide more steam and flow resistance and therefore pressure reduces. Flashing occurs in the core, reducing moderation and therefore reactor power. The converse occurs when the load is reduced, both to the opposite of the desired effect. This behavior is compensated by recirculation control. The jet pumps increase coolant mass flow proportionately to the required load increase. This reduces boiling, moderation and therefore reactor power increase. In a turbine trip or emergency load rejection, the turbine cannot accept reactor coolant. A dump system, the dry well, a concrete vessel surrounding the reactor vessel are provided. This is connected to a toroidal chamber below, the suppression pool, which contains water to condense the steam. As with the PWR, refueling must be off load at 12–18 month intervals, but it is a more complex operation because steam separators must be removed from the vessel head and the dry well and suppression pool must be flooded for radiological protection. Passive ECCS is actuated by low reactor pressure or high drywell pressure. If water level in the reactor vessel cannot be maintained, it is depressurized by discharge into the suppression pool, a core spray system is actuated, and the vessel is filled from below by a low-pressure flooding system. The steam turbines are very similar to those for the PWR, but subject to radioactivity in the steam. Most radioactivity is retained in the water recirculating in the reactor vessel, but a radioactive nitrogen isotope ^{16}N is carried to the turbine in the steam. However, its half-life is short and turbine maintenance can be commenced soon after shutdown.

PRESSURIZED HEAVY WATER (CANDU) REACTOR

In the 1950s, Canada developed a reactor design that optimized the use of national resources. Heavy water, for which production facilities had been developed during the Manhattan project, was chosen as moderator. This is the most efficient available and permits the use of natural uranium fuel, while avoiding the need for enrichment facilities. At that time, Canadian industry did not include heavy pressure vessel fabrication, so a pressure tube design was chosen. This is the Canadian Deuterium Uranium (CANDU) reactor, shown in Fig. 3. A 20-MW(e) demonstration plant was operated in 1962 and the first commercial plant of 200 MW(e) opened in 1968. CANDU is attractive to developing countries because it does not require an enrichment plant and has been bought by South Korea, China, Argentina, Romania, Pakistan and India, which has developed a national derivative. It is the world's third-most common design, with 29 plus 11 Indian versions in service with a total generating capacity of 22 GW. The core consists of a horizontal stainless steel cylindrical tank, called the Calandria, containing the heavy water moderator at atmospheric pressure. About 400 horizontal zircoloy tubes pass through it, each accommodating a pressure

Nat–Perf

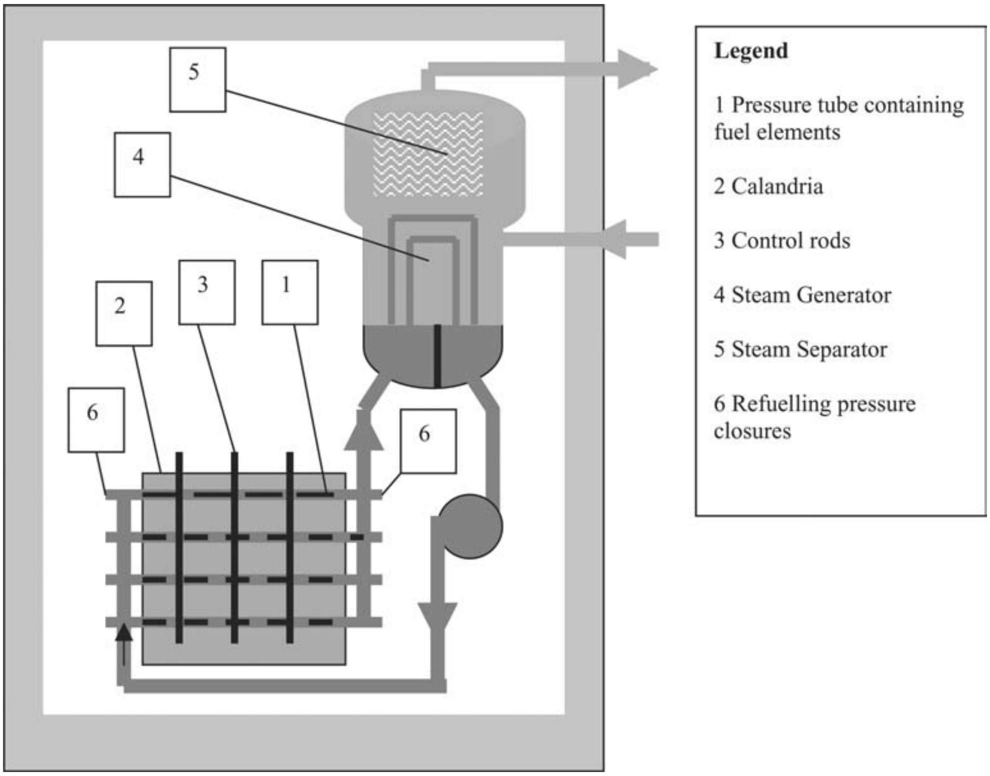


Fig. 3 Canadian deuterium uranium (CANDU) reactor.

tube, also zirconium, containing the fuel and heavy water coolant, at 90 bars. Coolant enters at 249°C and leaves at 293°C. The rods are natural U_2O pellets contained in zircolloy tubes that are 0.5 M long, 38 are connected to form a cylindrical bundle and 12 bundles are arranged end-to-end in the pressure tube. The pressure tubes are connected via headers to steam generators similar to those used in a PWR. The shell side carries light water and generates steam at 251°C and 40 bars. To minimize the effect of a LOCA, pressure tubes from the left and right sides of the core are connected in separate cooling circuits. Control rods are vertical and arranged to pass between the horizontal core tubes. The system is rapid-acting to compensate a slightly positive void coefficient. A secondary shutdown system injects liquid neutron absorbers into the Calandria. There is also a facility to dump heavy water from the Calandria in an emergency. Refueling is continuous on load and is managed to keep reactivity near constant, avoiding the requirement for the degradable poisons used in light water reactors and maximizing fuel utilization. Identical refueling machines are pressurized and connect to opposite ends of the selected pressure tube. Pressure closures on the tubes are opened to the refueling machines, maintaining coolant pressure. One machine injects new fuel, pushing the spent fuel into the other. The connections are then pressure closed and the machines are depressurized and moved to handling systems, which transfer the spent fuel to storage.

The exceptional neutron economy of CANDU was chosen to permit the use of natural uranium fuel. It is also sufficient to fission plutonium, resulting in once-through use of that produced by transmutation of ^{238}U during normal operation or enabling the use of light water reactor spent fuel or MOX fuel from decommissioned weapons. The Girdler Sulphide heavy water production process, used to establish the CANDU program, has now been discontinued because of its energy costs and toxicity, and an alternative is being investigated. CANDU has been the subject of controversy regarding the proliferation of nuclear weapons materials because of its ability to precisely manage plutonium production through continuous refueling. Future CANDU reactors will use light water coolant. This reduces neutron economy, necessitating a slightly enriched fuel, but also reduces the quantities of heavy water required and the system first cost.

GAS-COOLED REACTORS

Following the McMahon act, the United Kingdom developed independent nuclear weapons. Plutonium production reactors similar to the Hanford designs were built at Windscale. They differed in that they were cooled by air, forced through the fuel channels by fans, and exhausted to the environment via a system of filters, an arrangement that resulted in the first major fission product release from a civil plant. Early U.K. reactors were very efficient in breeding plutonium. The trend of natural uranium, graphite-moderated, gas-cooled reactors was continued with the Calder Hall design, which used closed circuit carbon dioxide cooling. To provide sufficient heat transfer, powerful gas circulators and a coolant pressure of about seven bars, requiring a steel pressure vessel, was used. Although these reactors were primarily for military plutonium production, reactor heat was recovered by external boilers to drive 60-MW(e) turbine generators. When commissioned in 1956, Calder Hall was celebrated as the world's first nuclear power station. The type is known as Magnox because of the magnesium alloy fuel cladding. It was continually evolved through 16 reactors of up to 600 MW(e) until 1970 and single examples were exported to Japan and Italy. Because of increasing size and gas pressure of up to 25 bars, a prestressed concrete pressure vessel, combining pressure containment and biological shielding, which also enclosed the boilers, was used for the last four to be built. However, Magnox cladding limited coolant temperature to about 400°C, fuel burnup to 9.0 GW-days/ton, and efficiency to 30%. The final Magnox design was further developed into the advanced gas cooled reactor (AGR) to overcome these limitations. The essential change was the use of stainless steel cladding, which allowed coolant temperatures of 700°C, thereby increasing energy efficiency to 40%. The lower neutron efficiency of stainless steel dictated a 2% enriched uranium oxide fuel. Burnup rates of 27 GW days/ton are now achieved, and 12×660 MW(e) AGRs of three significantly different designs were commissioned from 1976 to 1988. The policy of developing three parallel designs dissipated resources and included one which suffered 100% program overrun. Although the AGR concept is attractive, particularly with regard to safety and efficiency and includes on-load refueling, cost and time overruns on two of the designs resulted in its rejection in favor of the PWR before a previously ordered second generation AGR of the third design proved to be entirely successful and cost competitive. The only commercial gas-cooled nuclear power plant built in the United States was Fort St. Vrain in Colorado. It was a 330-MW, high-temperature helium cooled reactor (HTGR) first operated in 1976. A graphite moderator and particle fuel similar to Triso (described below) embedded in graphite blocks were used. It was 38% efficient, achieved rated load, periods of high availability, and demonstrated safe operation during interruption of coolant, but it was rendered uneconomic by persistent development problems unrelated to concept feasibility, and it shut down in 1989.

FAST BREEDER REACTORS

All of the reactors described above use thermal neutrons and burn only the small fissile component of natural or slightly enriched uranium. From the beginning of nuclear power, there was concern about the inefficient use of limited uranium resources. Fast breeder reactors (FBRs), which simultaneously generate electricity and, by progressively producing (breeding) and using plutonium, increase the energy extracted from uranium by a factor of more than 50, became the primary objective. Plutonium economies involving FBRs and fuel reprocessing began. In the United States, experimental fast reactors were operated from 1951 and the world's first commercial plant of 94 MW(e), named after Enrico Fermi, began in 1963. The latter was subject to core damage and other difficulties and was decommissioned in 1972. FBRs were discontinued by the Carter Administration because of concerns about proliferation of nuclear materials. At Dounreay in Scotland, a 14-MW(e) demonstration reactor was operated from 1959 to 1977 and a 250-MW(e) commercial prototype from 1976 to 1994 until government funding was withdrawn. France started later, but its 200-MW(e) Phoenix reactor was commissioned in 1973 and is still operating, now primarily for fuel cycle research. The 1200-MW(e) Super Phoenix, the largest fast reactor built to date, operated beginning in 1984, but shut down in 1997. In these countries, the availability of relatively low cost uranium fuels and the cessation of the growth of nuclear power removed the immediate need for fast breeder development. In the longer term, FBR and pyrometallurgical reprocessing (see nuclear fuel cycle) offer much higher utilization of uranium and the ability to burn transuranic wastes, which are currently the main long-term disposal problem. In the Soviet Union, a fast reactor program included the BN-600 power plus seawater desalination plant and design of reactors of up to 1600 MW(e). The former remains the most reliable Russian nuclear plant and, apparently, the world's most successful FBR. Development continues in Japan, which operated a 280-MW(e) fast reactor for about 2 years in the mid-1990s before it was shut down because a sodium coolant fire; and in India, which has had an ongoing program from 1985. The FBR concept excludes a moderator. The core is as compact as possible and power density is exceptionally high, of the order of 400 MW/m^3, or four times that of a PWR. Coolant with low neutron capture and very high heat transfer is required and core geometry must be precisely maintained to ensure uniform cooling. So far, liquid metal coolants, predominantly sodium, have been used with an outlet temperature of 600°C, yielding a thermal efficiency of 40%. Sodium has the advantage of boiling at about 1000°C and therefore the reactor operates at atmospheric pressure, giving very high heat transfer and being noncorrosive. Its disadvantages are that it becomes radioactive and reacts vigorously with water, necessitating an intermediate loop to exclude the possibility of water ingress from the steam turbine circuit into the reactor coolant. Helium cooling, which will increase outlet temperatures and therefore thermal efficiency, has been proposed. Fig. 4 shows a sodium

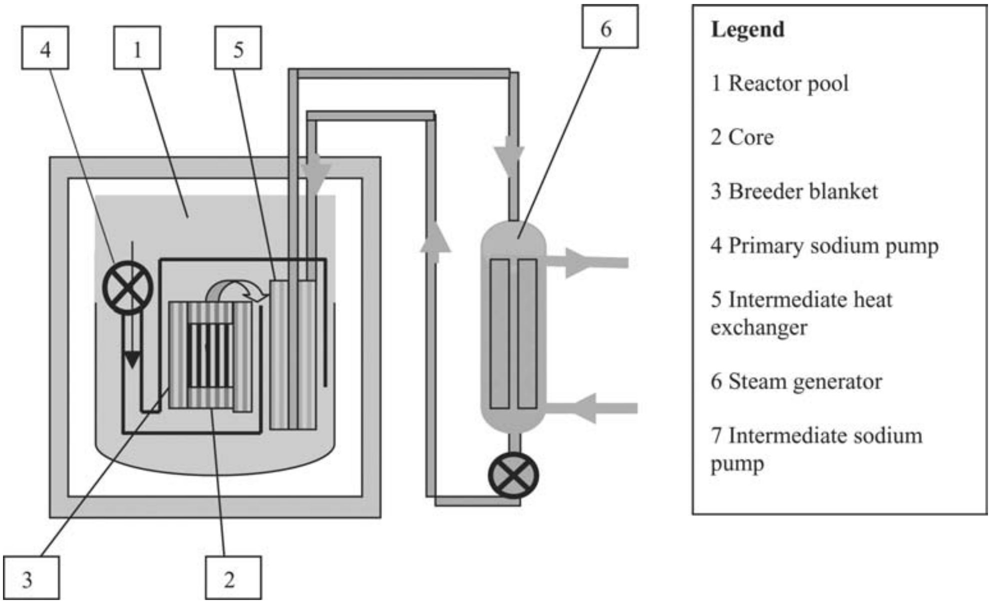


Fig. 4 Fast breeder reactor.

cooled pool type, typical of present FBRs. Phoenix, Super Phoenix, BN-600, and the U.S. advanced liquid metal reactor (ALMR) all used this concept. Other U.S. and Japanese designs used the alternative loop concept in which the intermediate heat exchanger is contained in a separate vessel. The pool is more compact and less sensitive to sodium leaks. The following description applies to Super Phoenix. The core consists of 360 stainless steel fuel assemblies, which are hexagonal to obtain the closest packing and 180-mm across flats. Each contains 200 8-mm diameter stainless clad rods of mixed, 20% enriched uranium and plutonium. It is surrounded by a breeder blanket comprising rods of fertile, depleted uranium, producing ^{239}Pu. These are arranged above and below the fuel in the core assemblies, and in 200 similar assemblies around the core circumference. The whole is surrounded by about 1000 stainless steel neutron reflector shields of similar dimensions. All of the above, supported by a structure, the diagrid, and the primary circuit, comprising sodium-to-sodium heat exchangers and pumps, are enclosed in a stainless steel tank about 60 mm thick. It is provided with a core catcher shaped to distribute molten fuel, preventing the formation of a critical mass in the event of a meltdown. The tank is surrounded by a secondary containment vessel designed to contain the coolant in the event of a breach. "Cold" primary sodium at about 390°C is pumped into radial slots at the bottom of the assemblies and flows upwards into a plenum at about 550°C, downward through shell side over the tubes of the intermediate heat exchangers (IHX), and exits into the volume below the plenum. Intermediate sodium enters the top of the IHX at about 345°C, is piped to the bottom, then passes upwards through the tubes and to the steam generator at 525°C. The steam generator is a once-through design. Feed water at about 200 bars is pumped upward through helically wound heat exchange tubes, in which it is boiled and superheated, leaving at 490°C. Refueling is at 2-year intervals, and limited by cladding irradiation damage, rather than fuel burnup. Load control, temperature, and burnup compensation are via stainless-steel-clad boron carbide control rods distributed across the core. A diverse shutdown rod system is also provided. A conventional steam turbine is used with reheat provided by a steam-to-steam reheater. Sodium/air heat exchangers are connected to the intermediate loops to provide decay heat removal if the steam turbine is unavailable.

FUTURE DEVELOPMENTS

The U.S. Department of Energy initiated the Generation IV program in 1999. It subdivides reactors into 4 categories: Generation 1 prototype commercial reactors developed in the 1950s and 1960s, modified, and enlarged from military applications; Generation II commercial reactors currently in service worldwide; Generation III developments of Generation II, some of which are already entering service; and Generation IV, which are advanced designs for application after 2020. Most Generation III reactors are light water, designed to be simpler, easier to operate, and less vulnerable to LOCA events. The safety philosophy is predicated on passive safety systems reliant on gravity, convection cooling or resistance to high temperature, and minimum human intervention in emergency conditions, rather than systems that must be relied upon to start and then be subject to human control and error. They are also designed for burnup of up to 65 GW-days/ton, using degradable poisons to extend fuel life, high availability, and 60-year operating lives. Another general objective is smaller modular reactors with passive safety features that enable a generating site to be progressively developed and are more suitable for small electrical power grids in developing countries. The first Generation III design, the 1300-MW ABR, has been deployed in Japan since 1997, with five now operating. Recirculating pumps and piping are enclosed within the reactor vessel, eliminating them as a cause of LOCA. Vessel safety is increased by the use of large forgings to minimize the number of welds and by eliminating large nozzles below the top of the core. Plant operation in a LOCA is fully automated. A further development is the economic and simplified boiling water reactor ESBR with a longer vessel to improve core cooling by natural circulation, and gravity-driven emergency cooling systems. A 1600-MW(e) European Pressurized Water Reactor (EPR), which is under construction in Finland for operation in 2009 and undergoing licensing for use in France, uses quadruple-redundant rather than passive emergency cooling systems. It can use 5% enriched or MOX fuel and is 36% efficient. Simplified, advanced PWRs with passive safety and enhanced reliability are NRC certified and ready for application in the United States. International reactor innovative and secure (IRIS) is an apt description of a 335-MW(e) PWR expected to be certified by 2015. All components, core, coolant pumps, steam generators, and pressurizer are enclosed in a single-pressure vessel, greatly reducing LOCA probability and power plant volume. Two gas-cooled reactor designs are also undergoing certification. The pebble bed modular reactor (PBMR) uses a radically different fuel comprising 60-mm diameter graphite balls, pebbles, filled with TRISO particles, 0.5-mm spheres of uranium or thorium oxide or carbide, each coated in successive layers of carbon, graphite, and a ceramic. The carbon allows for swelling due to fission products while the graphite is the moderator and containment. The pebbles therefore contain fuel, fission products as they are produced, two layers of containment, and moderator in a robust form resistant to temperatures over 2000°C. They also result in a very low core-power density. Helium is used as coolant and working fluid for a gas turbine with a core outlet temperature of 900°C, eliminating IHX. This concept has major safety advantages over other reactors. In a LOCA, fuel Doppler Effect increases neutron absorption by ^{238}U, slowing the reaction and achieving a steady state at which temperatures, acceptable because of the fuel design, are maintained by passive air-cooling. Emergency cooling is not required and efficiency is high because of the high temperatures in the gas turbine cycle, 45%–50%, depending on the specific reactor design. Refueling can be continuous, enabling periodic remote pebble inspection, and it constantly moves them through the core to increase burnup. The PBMR was devised and a 14-MW(e) prototype successfully operated in Germany for 20 years, including a specific demonstration of the inherent safety of the design. It was abandoned following Chernobyl, but is now being developed in South Africa and China. South Africa has approved the construction of a 110-MW(e) prototype in 2007. The concept is shown in Fig. 5. The pebble bed core is surrounded by a graphite block neutron reflector. A

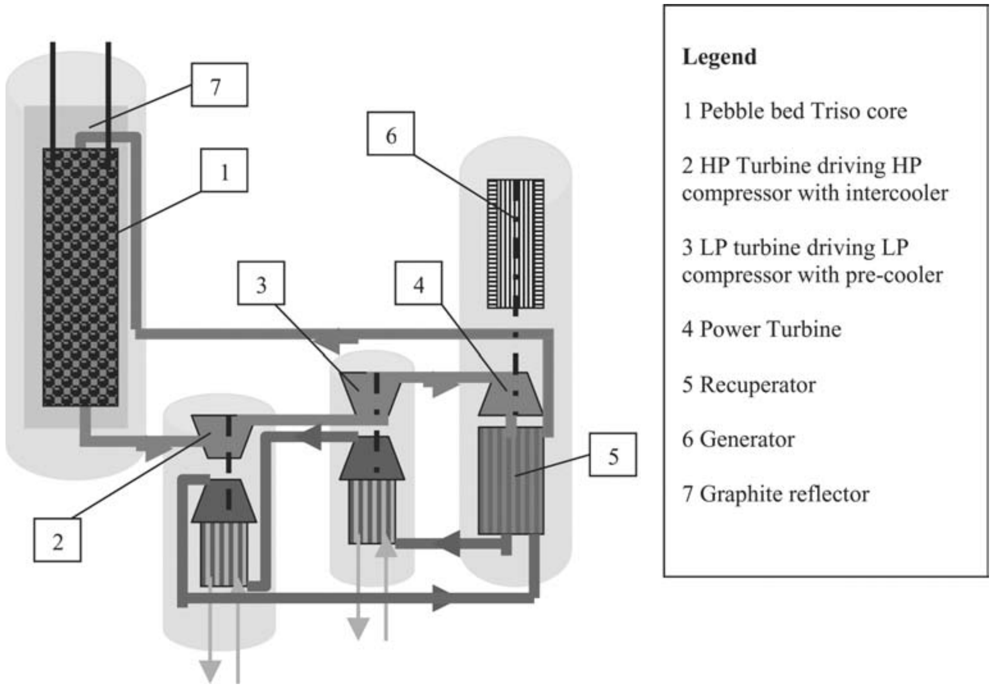


Fig. 5 Pebble bed modular reactor (PBMR).

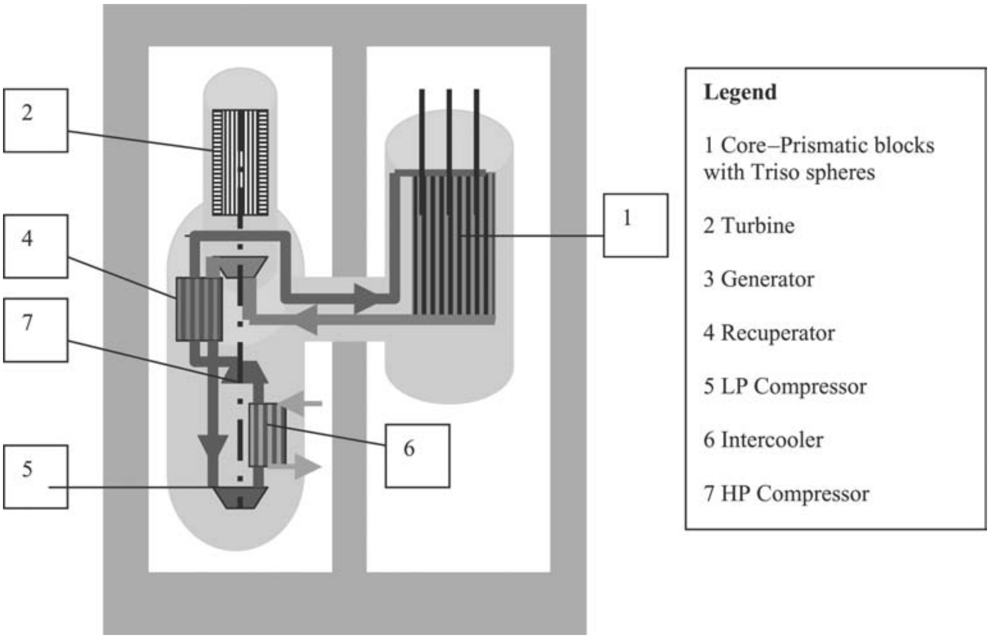


Fig. 6 Gas turbine modular high temperature reactor GT-MHR.

closed Brayton cycle is used with a recuperator and three gas turbines, using magnetic bearings, to drive two stages of gas compression and the generator. Containment is dependent on the integrity of the fuel and the reactor vessel. There is no secondary containment or emergency cooling system. China operates a small prototype PBMR and plans a 200-MW production design and large-scale adoption by 2020. The second design being certified is the 285-MW(e) gas turbine modular helium reactor, GT-MHR, shown in Fig. 6, developed jointly by the United States and Russia. It was initially planned to burn surplus Russian weapons grade plutonium, but can also use up to 20% enriched uranium fuel. Fuel configuration is Triso particles contained in hexagonal graphite blocks. Its safety philosophy is high temperature passive cooling, as described above. It is a compact, modular design with vessels for both the reactor and for the power system. The latter comprises the gas turbine generator and two gas compressors on a single shaft and the heat recovery system. Secondary containment is provided. Generation IV developments of both water and gas-cooled reactors are in progress. Supercritical water cooled reactors (SCWRs) are a combination of supercritical boiler and steam turbine technology, well proven in fossil fired steam plant, and the BWR. The reactor coolant is single phase; therefore, jet and recirculating pumps are eliminated. The specific enthalpy of coolant is increased therefore the reactor is more compact. The coolant outlet temperature is increased and efficiency is over 40%. The use of a fast neutron core further increasing compactness is being considered. Extensive computer simulations of LOCA show that safe operating temperatures for stainless-steel cladding will not be exceeded. Helium-cooled fast reactors based on GT-MHR and a very high temperature reactor (VHTR) using a thermal core with 1000°C helium output temperatures are being developed.

FUSION REACTORS

Controlled fusion reactions for power production have huge potential as inexhaustible energy sources that do not produce long-lasting radioactive waste. This potential has justified fusion research in Europe, the United States, and Russia for half a century and the results achieved now justify continuation for another 30 years in the ITER project now being built at Cardarache in France. However, the engineering difficulties of power production—i.e., stable containment of plasma at 100 Million °K (Kelvin) in a practical reliable system—are probably the most formidable ever faced and success is uncertain. Fusion reactions are discussed in "Nuclear Technology." A brief summary of the history of fusion research and the objectives and concepts of ITER is as follows. The first plasma was created in the United Kingdom in 1947, with the evolution of toroidal containment devices occurring there; ZETA was commissioned in 1954 and JET in 1983, resulting in the highest fusion power so far achieved—16 MW for about 1 s in 1997. The Soviet Tokamak provided a major development in containment technology through its combination of magnetic fields produced by external electromagnets and the helical field component due to plasma current, and achieved plasma temperatures of 10 million °K. A device of this type was built in the

United States in 1980; and in 1998, a Japanese version achieved a ratio of fusion power to input power (Q) of 1.25, the highest yet achieved. Construction of ITER, intended to prove feasibility of fusion as an energy source, commenced in 2005. Its objectives are to produce 500 MW of power for 500 s and a sustained Q of 5. The major developments required are plasma stability, materials for the vacuum vessel, which is subject to intense neutron and thermal radiation, and breeding tritium fuel from lithium in an outer blanket. Energy will be extracted by neutron bombardment of a primary coolant, but in ITER, it will dumped as waste. It is expected that a demonstration power plant will be developed as ITER proceeds, with the objective of power production in 2040.

CONCLUSION

The most popular nuclear power plants, light water reactors, which originated in military applications, have undergone extensive development and have been built in large numbers, creating an application momentum that suppressed inherently safer and potentially economically competitive designs. Gas-cooled reactors with great potential were built in the United Kingdom, Germany, and the United States, but were unsuccessful because of delays and development problems. The renewed emphasis on safety has further enhanced that of light water reactors via development of passive or more complex safety systems, and also renewed interest in the inherent passive safety of gas-cooled designs. This and the trend toward smaller reactor sizes, which are more suitable for developing nations and the progressive development of generation inventories, is establishing a range of gas- and water-cooled reactors that are well suited to the increasingly accepted, major role that nuclear power generation must play in satisfying the world's energy demands, while minimizing CO_2 emissions. Uranium availability permits large-scale use of nuclear power with a once-through fuel cycle, but, in the longer term, the fast reactor remains attractive as a means of minimizing high-level waste. ITER promises unlimited energy with minimum pollution, but also formidable development problems.

BIBLIOGRAPHY

1. A Roadmap to Deploy new Nuclear Power Plants in the United States by 2010, United States Department of Energy Office of Nuclear Energy, Science and Technology and its Nuclear Energy Research Advisory Committee Subcommittee on Generation IV Technology Planning (October 31, 2001) website http://www.nuclear.gov/ (accessed August 2006).
2. El Wakil, M.M. Thermal fission reactors and power plants. *Power plant Technology* 1st Ed., McGraw Hill: New York, U.S.A., 1984; 395–452, [Chapter 10].
3. El Wakil, M.M. Fast breeder reactors and power plants. *Power plant Technology* 1st Ed., McGraw Hill: New York, U.S.A., 1984; 453–498, [Chapter 11].
4. Central Electricity Generating Board Nuclear Power Generation In *Modern Power Station Practice*, Vol. 8; Pergamon Press: U.K., 1971.
5. Oka, Y.; Koshizuka, S. Supercritcal pressure, once-through cycle light water reactor concept. J. Nucl. Sci. Technol. **2001**, *38* (12), 1081–1089.
6. Agnew, H. Gas—cooled nuclear power reactors. Sci. Am. **1981**, *244* (6).
7. Knief, A. *Nuclear Energy Technology, Theory and Practice of Commercial Nuclear Power*; McGraw Hill: U.S.A., 1981.
8. Main Features of ITER; http://www.iter.org/FeaturesD.htm (accessed August 2006).
9. Kurpitz, J. Mourgov, V. The role of small and medium-sized reactors. Twenty Third Annual International Symposium; The Uranium Institute: London, 1998.
10. Wilson, R. The changing need for a breeder reactor. Twenty Fourth Annual International Symposium; The Uranium Institute: London, 1999.

Nat–Perf

Nuclear Energy: Technology

Michael Burke Wood
Ministry of Energy, Al Dasmah, Kuwait

Abstract
This entry summarizes the history of nuclear technology, presents essential concepts of atomic structure, binding energy, and the three types of nuclear reaction—radioactivity, fission, and fusion. General implications for reactor design, control, and operation are given. Nuclear safety, accidents, and their implications and the effects of radioactivity are discussed.

INTRODUCTION

Nuclear technology and today's power reactors originated in weapons programs during and immediately after World War II. The various choices of moderator and coolant then made for reasons other than optimization of civilian nuclear power plants determined safety characteristics, which are the subject of continuing development. The very severe responsibility for public safety has resulted in a continuously developing and in depth safety philosophy unique in all industry. The high-profile accidents that have occurred and the operating record following these accidents have provided clear comments on its success. This article summarizes the fascinating history of nuclear technology—its essential principles, safety issues, and the significance of the accidents. Nuclear power plants and fuel cycle are discussed in companion articles.

HISTORY

In about 400 B.C., the Greek philosopher Democritus suggested that matter is composed of common fundamental particles, which he named "atomos," (literally, indivisible). Scientific justification came in the early 19th century, when from experiments Dalton deduced that elements combine in fixed ratios, confirming the existence of a fundamental unit of mass for each—the atom. After half a century of controversy, Avogadro's hypotheses on combination into molecules were accepted. Faraday's electrolysis experiments demonstrated that electrical forces exist within molecules and the fundamental electric charge associated with ionization. Identification of subatomic particles began in 1897 when, from deflection of cathode rays by electric fields, Thompson identified electrons as negatively charged particles with about 1/2000th of the mass of a hydrogen ion. The discovery of radioactivity of uranium, long known as a coloring agent for glass and ceramics, by Becquerel and of radium by Pierre and Marie Curie were further evidence that the atom was divisible. Rutherford identified alpha radiation as helium nuclei and beta radiation as electrons. Villard showed gamma to be short wavelength electromagnetic radiation. In 1911, Rutherford observed occasional scattering of alpha particles by very thin gold foil, and reasoned that most of an atom must be empty space with its mass concentrated in a tiny nucleus. Soon afterwards, Bohr proposed his model of atomic structure with electrons in orbits (shells), which explained atomic stability and quantized energy emissions.

In 1918, Rutherford found that hydrogen was produced when nitrogen was bombarded with alpha particles and deduced that the hydrogen ion must be a fundamental component of all nuclei, which he named the proton. Fourteen years later, Chadwick identified the neutron. These discoveries were all made with relatively simple experiments, but further progress required high energy.

In 1932, Cockcroft and Walton broke nuclei using protons accelerated in a high-voltage vacuum tube. Lawrence began the trend to high-energy research machines with his cyclotron, in which charged particles were accelerated in a spiral. Using one, Curie and Joliot produced artificial, beta-radioactive nuclei by alpha bombardment. In 1934, Fermi used neutron bombardment, eliminating repulsive forces and so obtained higher collision energies. He experimented with several elements, including uranium, and produced new radioactive nuclei that he could not identify. Rutherford doubted that energy production from nuclear processes was feasible, but in 1935 both Slizard and Joliot predicted chain reactions releasing large amounts of energy. Supporting evidence came in 1938 when Hahn and Straussman repeated Fermi's experiment and showed that uranium had fissioned into two stable nuclei. Frisch and Meitner interpreted these results using theoretical models of

Keywords: Atomic structure; Radioactivity; Fission; Fusion; Reactor design; Safety; Accidents; Radiation.

nuclear binding energy and confirmed that fission was accompanied by a large release of energy.

The two possibilities—a controlled chain reaction in a distributed mass, a reactor; and a concentrated mass reacting exponentially, a bomb—were now foreseen. Bohr, working with Wheeler, identified the major difficulty of the latter—only the U235 isotope, 0.7% of natural uranium, underwent fission, and the rest, U238, absorbed neutrons and inhibited the reaction. A natural uranium bomb was therefore impossible and separation of the two isotopes was essential. This was known to be extremely difficult because of identical chemical and near-identical physical properties and because the quantity of U235 required was uncertain.

Civil nuclear power was first suggested by Werner Heisenberg in a 1939 paper in which he proposed that a low level of enrichment would be sufficient and the use of heavy water or carbon moderators, the former being optimum. A group of researchers in Paris reached the same conclusions, and both attempted to obtain heavy water, which was only produced in Norway as a by-product of ammonia synthesis. The Germans were outwitted and the world's only supplies were transported to France. As the Germans invaded, members of the Paris group escaped to England with the heavy water. In 1939, on the eve of Word War II, physicists in the United States, led by Slizard, were concerned by the German lead in nuclear physics, indicated by Frisch and Meitner's work, and that Nazi Germany would build a bomb. They drafted a letter to President Roosevelt, advising him of the possibility and persuaded Albert Einstein to sign it to add credibility.

Because the United States was not at war and due to uncertainty about the feasibility of the bomb, the letter resulted in only small increases of funding for nuclear research. Frish, now in Britain, reported to a government committee that only a few kilograms of U235 were required for a bomb. The Committee's findings, including the conclusion that isotope separation by gaseous diffusion was feasible, were reported to the United States. Bohr and Wheeler predicted that an artificial nucleus of atomic weight 239 would also be fissionable, and in the United States, Seaborg produced the element, plutonium, by deuteron bombardment of U238 in a cyclotron. This offered the second bomb-making route of plutonium production in a reactor and relatively easy separation because of the difference in chemical properties. The British report was accepted, and just before Pearl Harbor, Roosevelt initiated the Manhattan Project for producing nuclear weapons.

The Manhattan Project's first major achievement, in December of 1942, was Fermi's Chicago Pile-1. This was literally a 1300-ton flattened spheroidal pile of graphite moderator blocks, with some including pellets of natural uranium with cadmium control rods. It was uncooled and operated only for a few minutes at negligible power, but proved the feasibility of a continuous chain reaction. Under military administration headed by General Groves and scientific direction by Oppenhiemer, the Manhattan Project accelerated into one of the largest and most successful scientific and industrial co-operations in history. The British program was absorbed and the project rationalized into three main centers—plutonium production at Hanford, uranium enrichment at Oak Ridge, and weapons research and design at Los Alamos. Enrichment was primarily by gas diffusion. A full-scale alternative plant using a form of mass spectrometer called a Calutron, invented by Lawrence, was also used but was less efficient. Both uranium and plutonium bombs were developed and used. Because the latter required an uncertain implosion trigger mechanism, the design was tested in New Mexico before use at Nagasaki in 1945. Spying aided the Soviet nuclear weapons program. This influenced the 1946 McMahon Act, which placed all U.S. nuclear technology under the control of the civilian U.S. Atomic Energy Commission, the forerunner of the Nuclear Regulatory Commission (NRC), and banned all international technology exchange. Britain and Canada were forced to revert to independent programs. Despite the concerns, German wartime nuclear technology was unsuccessful—the exodus of physicists due to anti-Semitism, lack of heavy water and the impurity of available graphite, the alternative moderator, all impeded it—but uranium was available and the best isotope separation technique, the gas centrifuge, was invented. It remains controversial whether Heisenberg and his cow-orkers deliberately impeded Nazi nuclear weapons or grossly overestimated the fissile material required, therefore considering them infeasible.

The Manhattan Project, the McMahon Act, and cold war military requirements influence nuclear technology to the present day. Canada followed the only purely civil nuclear power program. The CANDU reactor was defined by the lack of uranium enrichment and the availability of heavy water. In the United States, with enrichment and plutonium production already available, compactness for submarine propulsion dominated reactor design, resulting in the Pressurized Water Reactor (PWR) and its civilian derivative the Boiling Water Reactor (BWR). In the United Kingdom, France and Russia, all initially without these facilities, graphite-moderated natural uranium reactors with gas or water-cooling were chosen for plutonium production. The United Kingdom persevered with gas graphite reactors because of higher thermal efficiency and inherent safety until the 1980s, but all of these countries eventually selected versions of the PWR because of the large body of supporting experience and the supposedly lower capital cost. Early expectations of the domination of nuclear electricity generation and a consequent shortage of uranium led to fast breeder reactor development programs. These terminated in the United States in 1977 because of concerns of proliferation of nuclear materials and in other western countries in the aftermath of the Three Mile Island and Chernobyl accidents. Although the former increased confidence that the consequences of a very serious accident could be confined to the plant and the latter had no relevance to any reactor in the West, the effect of these accidents on public opinion and politics effectively froze development and construction of nuclear plants in the West for two decades. Concern about global warming and alternative energy supply is now stimulating a renaissance of both.

ATOMIC STRUCTURE

Atoms comprise a central nucleus made up of nucleons; positively charged protons and neutral neutrons, which are most of the mass, surrounded by negatively charged electrons. Atomic number Z is the number of protons. In a neutral atom, it is also the number of electrons. Mass number A is equal to the number of nucleons. An atom is represented $_ZX^A$, where X is the chemical symbol for the element.

The atomic number determines the grouping and bonding of electrons and how they are exchanged or shared between atoms, and therefore the chemical properties. Mass number determines nuclear and magnetic properties and weight. Isotopes have the same atomic but different mass numbers. They have the same chemical properties, slightly different physical properties, and different nuclear properties.

BINDING ENERGY

When a nucleus disintegrates, energy is absorbed. This is equal to the work done to move the nucleons apart against the binding forces, into a state of less stability and higher potential energy. Conversely, when a nucleus is formed, its nucleons move into a state of lowest potential energy under the binding forces and energy is released. In any reaction, the net energy release is the difference between these two effects. In chemical reactions it is relatively small. Only electron shells participate and energies are a few electron volts (EV) per product molecule. In nuclear reactions where energy releases per nucleus are hundreds of millions of EV, it is prodigious—the reason for the existence of nuclear technology.

The EV is an energy unit used in physics that represents the energy gained by an electron passing through a potential difference of 1 V, equal to 1.6×10^{-19} J. Energy and mass have an equivalence given by $E=mc^2$. In a reaction, mass-energy is conserved; thus, when energy is released, mass is reduced.

A mass deficit of 1 kg results in 9×10^{16} J of energy release. In the U235 fission reaction described below, the mass deficit is about 0.09%, and energy release about 200 MeV per U235 nucleus, 8×10^{13} J, 1 GW-day of thermal energy per kg of U235 fissioned completely. This is equivalent to firing 3000 ton of coal or 13,000 barrels of oil. Complete fission cannot be achieved. Fuel burn-up in present reactors gives about 2% of the above energy release and hydrocarbon fuel equivalence. Nuclear binding force is the resultant of the attractive strong force and the repulsive electrostatic force. The former acts only between adjacent nucleons and may be visualized as surface tension in a droplet of liquid representing the nucleus. The latter acts between all protons and follows an inverse square law. Because of these characteristics, binding force per nucleon depends on the size of the nucleus, i.e., mass number, increasing until 56, (iron) and then decreasing at higher mass numbers. Thus, energy is released both when light nuclei combine to form a heavier, more stable nucleus in a fusion reaction and when a heavy nucleus divides into two lighter, more stable nuclei in a fission reaction. A further consequence is that stability of large nuclei depends on excess neutrons, which increase the attractive forces. $_{92}U^{238}$, which is 99.3% of natural uranium, has 146 neutrons and is the most stable of its isotopes. $_{92}U^{235}$, which, neglecting a rare 3rd isotope is the remaining 0.7%, has 3 less and is the least stable.

RADIOACTIVITY

Radioactivity is spontaneous nuclear decay of unstable isotopes. It occurs in about 50 natural and all artificially produced isotopes and proceeds through a chain of decay reactions until a stable isotope is formed. For uranium 238, there are 19 decay stages to lead. The time at which an individual decay occurs is random, but statistically, rate of decay is proportional to the number of isotopes (N) and average life ($1/\lambda$). Thus, $dN/dt=-\lambda N$. Integrating gives $N_t=N_0e^{-\lambda t}$, hence time to decay completely is infinite, but half-life—the time for decay to half the original quantity—is constant for each isotope and given by $\log_e 2/\lambda$. After x half-lives, the fraction of isotopes remaining and therefore the level of radioactivity is reduced by $(1/2)^x$. Half-lives are proportional to stability and generally inversely proportional to intensity of radiation. Those of some naturally occurring isotopes are comparable with the age of the earth. Uranium 238 is 4.4×10^9 years; U235, 7×10^8 years; and Plutonium 239, 2.5×10^4 years. The fission products, strontium 90 and cesium 137, are about 30 years. There are three types of radioactivity. Alpha decay is emission of a helium nucleus and reduction of atomic number by 2 and mass number by 4. It applies to all nuclei of atoms with an atomic number greater than 84. An example in the decay chain of U238 is formation of Thorium:

$$_{92}U^{238}\rightarrow{}_{90}Th^{234}+{}_2He^4$$

Beta radiation is of two types. Beta $^-$ decay occurs when a nucleus has one excess neutron. This transforms into a proton, thus the atomic number is increased. The associated weak force carrier particle is released and immediately decays into an electron and an antineutrino. For example, thorium Beta decays to protactinium:

$${}_{90}Th^{234} \rightarrow {}_{91}Pa^{234} + {}_{-1}e^{0} \ + \ \text{antineutrino}$$

In Beta $^+$ decay, an excess proton transforms into a neutron and the atomic number is reduced. The weak force antiparticle decays to a positron and a neutrino. For example, beryllium decays to lithium:

$${}_{4}Be^{7} \rightarrow {}_{3}Li^{7} + {}_{1}e^{0} \ + \ \text{neutrino}$$

After alpha or beta decay, the nucleus is generally not in its lowest energy state. Gamma radiation, high-energy photons (electromagnetic radiation), is therefore emitted. X radiation is similar, but its origin is acceleration of electrons outside the nucleus. Its wavelength is longer and its energy is lower.

FISSION

Most heavy nuclei are fissionable. This can occur spontaneously or be induced by other particles, but only neutrons can produce a chain reaction in which new inducing particles are emitted for every one absorbed. Nuclei with odd numbers of protons are less stable and fissile, i.e., fissionable by thermal neutrons, defined as being in thermal equilibrium with the surrounding atoms in random motion, with energies of about 0.025 MeV. Other isotopes with even numbers of protons are only fissioned by fast neutrons—those with energies greater than 0.1 MeV. U235 is the only naturally occurring fissile isotope. U233, plutonium 239 and 241, produced by transmutation of thorium 232, uranium 238, and plutonium 240, respectively, and some other transuranic elements (those above uranium in the periodic table), are also fissile. Nuclei, which can produce fissile isotopes by transmutation, are called fertile.

An example of a U235 fission reaction resulting in a fission product pair of xenon and krypton is:

$${}_{92}U^{235} + {}_{0}N^{1} \rightarrow {}_{92}U^{236} \rightarrow {}_{54}Xe^{141} + {}_{38}Kr^{93} \ + \ 2{}_{0}N^{1}$$

Hundreds of other fission product pairs with mass numbers between 72 and 160, which preserve the 92 protons of the uranium nucleus, may be produced, depending on the energy of the inducing neutrons. The division of mass becomes more even as energy increases. About 85% of the energy release is kinetic energy of the fission products, which is immediately converted into heat. The balance is in gamma radiation and the neutrons emitted. The number of prompt neutrons—those released immediately on fission—may be 2 or 3, depending on division of mass, with an average of 2.47. Fission products are unstable and beta decay. They also emit high levels of associated gamma radiation, possibly for very long times, depending on the product pairs. The associated decay heat, about 6% of the total, would cause fuel overheating and must be continuously removed, even from a shutdown reactor. In some fission products, the energy released by beta decay exceeds the binding energy of one of its neutrons, which is then released. Mean half-life for this release is about 13 s. Such delayed neutrons are only about 0.6% of the total for U235 and 0.2% for Pu 239, but are essential for reactor control.

After fission, the neutrons produced may: (i) escape from the surface of the material; (ii) be scattered, i.e., collide with a nucleus, reducing their energy; (iii) be captured by another nucleus, causing it to become radioactive and in some cases undergo transmutation to a chemically different nucleus; or (iv) cause a further fission. The most important example of capture is that resulting in transmutation of U238 into plutonium after beta decay of U239 and neptunium 239:

$$\text{Capture,} \ {}_{92}U^{238} + {}_{0}N^{1} \rightarrow {}_{92}U^{239}, \ \text{Beta decay} \rightarrow {}_{93}Np^{239},$$

$$\text{Beta decay} \rightarrow {}_{94}Pu^{239}$$

CHAIN REACTION

The requirement for a chain reaction is that neutrons produced by fissions cause further fissions. Multiplication factor K is defined as (neutrons produced by fission—neutrons lost by escape—neutrons captured) per fission. If $K=1$, the system is critical and reaction will be sustained at constant rate. If $K<1$, it is subcritical and reaction will decrease exponentially. If >1, it is supercritical and reaction will increase exponentially. If criticality is achieved by prompt neutrons only, the system is prompt critical. For safe control, reactors are designed to be delayed critical—criticality is achieved only with the addition of delayed neutrons. Escape depends on the ratio of surface area to volume, which is inversely proportional to size. Therefore, for each fissile material and configuration, there is a critical size and critical mass below which the reaction will not occur. The probability of each of the other processes is expressed as a cross section, an effective target area surrounding the nucleus through which an incident neutron must pass for the specific process to occur. Large cross sections indicate probable events. The unit is the barn, 10^{-28} M^2—the area of a typical nucleus. Cross sections are determined experimentally and depend on nucleus, neutron energy, and process. Values for homogenous assemblies of the most common nuclear fuels are:

	Fast neutrons			Thermal neutrons		
	Scattering	Capture	Fission	Scattering	Capture	Fission
Natural uranium	6.6	0.14	0.30	8.0	3.5	4.2
U238	6.6	0.14	0.02	8.0	2.7	0
U235	5.3	0.09	1.20	6.0	110.0	580.0

Plutonium 239 is similar to U235 but has a higher thermal fission cross section. From the natural uranium results it can be seen that small quantities of U235 markedly increases fission cross section. This trend is continued by low levels of enrichment. The probability of scattering and capture by the other materials required in practical systems is also expressed in barns.

Fast neutrons are those newly produced by fission with energies >0.1 up to 12, with an average of 2 MeV. Their energy is partially reduced by scattering, but thereafter in U238, capture is very high at intermediate energies, known as the resonance region where peaks in cross section occur. Fast fission is improbable and few neutrons remain to be further reduced to thermal energies. In U235, capture is less and fast fission more probable. The possibilities of fast fission devices are therefore as follows: for a bomb, the requirement is the maximum rate of reaction and energy release before the fuel self destructs, therefore a maximum value of K. Fully enriched U235 or plutonium, to minimize capture, and triggering mechanisms using chemical explosives to assemble a prompt super-critical mass, are required. For a fast reactor, as in all reactors, the requirement is a sustained, controlled reaction, therefore $K=1$, delayed criticality, as described above. This can be achieved with 20%–30% enriched fuel. No moderator is required. Fast reactors are potentially advantageous because although fast fission cross sections are less than for thermal fission, each fission, particularly those of Pu239, produces more neutrons; therefore, the reaction is more efficient for transmutation either of U238 or of high level nuclear waste (as discussed under fuel cycle). Conversion ratio is defined as fertile nuclei converted / fissile nuclei consumed. For U238 to Pu239, a ratio >1, a breeder, is only possible in a fast reactor using a breeder blanket of fertile material surrounding the core. Most thermal reactors that have a ratio <1 are burners, however, a thermal reactor using thorium 232 to breed fissile U233 is possible.

U235 and natural uranium, and from the above discussion, enriched fuels have their highest fission cross sections for thermal neutrons. However, in homogenous assemblies of natural and slightly enriched uranium, most neutrons are captured before they can be scattered to thermal energies. This problem is overcome by arrangement of the fuel in a matrix of small—typically about 0.4″ in diameter—fuel rods to allow fast neutrons to escape from them before capture. The space between rods is occupied by a moderator chosen to scatter neutrons to thermal energies before they re-enter fuel with a much increased probability of causing fission. An optimum moderator is a light nucleus with a low capture cross section. The best is heavy water, deuterium oxide, which is 0.015% of normal water. Because of the difficulty and cost of its separation, graphite is also used. Only these two moderators allow $K=1$ to be achieved with natural uranium fuel. "Light" normal water is the most commonly used but capture is sufficiently high to require the use of a 2%–3% enriched fuel.

REACTOR DESIGN

Many reactor designs using various fuels have been and are being developed. The most important are described in the entry "Nuclear Power Plants." Some fundamentals are discussed here. The core, the central region in which the reaction takes place, comprises fuel assemblies, moderator (in thermal reactors), and provisions for heat removal and control. It is designed to produce as uniform a distribution of reactivity and heat release as possible. Extending the moderator beyond the core as a reflector to scatter neutrons back into the core reduces major variations. Design is limited by the interaction of materials properties and reactor cooling. Where reactivity is low as in natural uranium reactors, moderator, coolant, and rod materials are chosen for minimum capture. Material properties then dictate rod-operating temperature, hence coolant outlet temperature. This defines the temperature differential over which the heat engine driving the electricity generator can operate and therefore its efficiency.

In the United Kingdom, gas cooling was dictated by the choice of natural uranium fuel, but it had inherent safety advantages by avoiding the sudden drop in heat transfer when film boiling developed in a water cooled reactor. Heat transfer and core power density are low, dictating a large core. Rods were magnesium alloy, limiting outlet temperature to 250°C, efficiency to 32%, and power density to 1 MW/M^3. Later, enriched fuel relaxed the capture constraint, allowing stainless steel rods and increases to 650°C, 42%, and 2 MW/M^3, respectively.

In the United States, enriched fuel was available and water cooling was chosen to provide high heat transfer, enabling a very compact core for submarine applications. To ensure this, boiling must be suppressed; therefore operating temperature is limited by saturation temperature at the highest pressure for which the pressure vessel can be designed. Rod material properties are therefore not a constraint, and a relatively low strength zirconium can be used to increase neutron economy. In current PWRs, the pressure vessel is limited to about 150 bars, coolant outlet to 320°C, and efficiency to 32%. Power densities are up to 100 MW/M^3. In a fast reactor, no moderator is required, fuel rods are very closely spaced, and power density is exceptionally high—400 MW/M^3. Coolant with low capture and very high heat transfer are required and core geometry must be precisely maintained to ensure uniform cooling. So far, liquid metals, predominantly sodium, have been used with outlet temperatures of 600°C and efficiency 40%. Helium, which will increase both parameters, has been proposed. Some early plutonium production reactors used open circuit cooling, i.e., the coolant was discharged to the environment. Closed circuit cooling is now universal, and in most reactor types, a heat exchanger or steam generator is provided between the reactor coolant in a primary circuit and a different working fluid, normally steam, in a secondary circuit for the heat engine. This complexity and the loss of efficiency due to temperature differentials in the heat exchanger is justified because at the temperatures involved, the steam turbine is the most efficient heat engine, and its radioactive contamination is avoided. In BWRs and high-temperature gas-cooled reactors, reactor coolant is the working fluid.

Fuel pellets are typically 10 mm in diameter, inserted into rods—thin wall, seal welded, metal tubes. Rods retain fission products and alpha and beta radiation and are designed to withstand internal pressure due to gaseous fission products and fuel distortion. They are made up into precisely spaced arrays called fuel assembles, which are removed and replaced during refueling. As reactor operation proceeds, fissile material, including some of that newly produced by transmutation, is burned and fission products, which absorb neutrons, build up, and therefore reactivity decreases. Gaseous products including krypton and xenon increase internal pressure. High neutron flux damages the metallurgical structure of the rods. It is advantageous to refuel on load to manage reactivity in the core and optimize burn-up and transmutation in the fuel. This is possible in some types of reactors, but not in light water reactors because of the difficulty of providing and sealing suitable openings in the pressure vessel. Shutdown for refueling is a major operation and a large proportion of the core is therefore replaced each time. Burn-up is measured in MW days of heat release per ton of fuel. Present reactors burn most of the fissile material in the fuel. Most generation II reactors (see nuclear power plants) achieve about 2×10^4 MW/D/T using a 2%–3% enriched fuel. However, most of the potentially fissile U238 and plutonium remains unused. This is discussed in "Nuclear Fuel Cycle."

Reactor Control is required to (i) ensure safe shutdown (SCRAM) under any condition, (ii) compensate for changes in reactivity (iii) regulate reactor power to follow electrical load. "SCRAM": is universal industry usage and is an acronym for Safety Control Rod Axe Man, dating to the Chicago Pile, with obvious implications. Methods of control are rods of neutron absorbing materials such as boron or cadmium, which are inserted into, or withdrawn from the core, supplemented, in water reactors, by liquid absorbers added to the coolant. Shutdown rods contain enough absorber to terminate the reaction under any conceivable condition and have fail -safe operation. Excess fuel is provided and compensates for the reduction in reactivity discussed above, when additional control rods, shim rods, are withdrawn. In PWRs, these are supplemented by decreasing the concentration of chemical shim—usually boric acid dissolved in the coolant. Increased capture also affects short-term operation. Some fission products with very high capture cross sections, particularly xenon 135, concentrate during shutdown or periods of load reduction because of a longer half-life than other fission products. This is xenon poisoning and it requires a further increment of additional fuel, compensated by more control rods to ensure a restart. For ease of control, total reactivity is arranged to be delayed critical, i.e., regulating rods have only a small effect and power increases occur slowly due to delayed neutrons. Control must also compensate Fuel Doppler effect, the increase of capture by U238 with temperature; Moderator temperature coefficient, an increase in scattering with temperature; and Void coefficient, an increase in reactivity due to vapor formation in a two-phase coolant. The combined increase in reactor power with temperature is the power coefficient. In light water reactors, all oppose a power increase and therefore the power coefficient is negative, increasing safety. This is not the case in the Chernobyl RMBK reactor.

FUSION

Fusion is the source of energy in stars. The essential process is fusion of 4 protons to form a helium nucleus; $4\ p\rightarrow He^4+2e+\text{energy}$. This may be by either of two sequences of reactions. The proton–proton chain, successive formation of deuterium, helium 3, and helium 4, predominates in smaller stars. The carbon cycle, successive reactions of protons with of nuclei of nitrogen, carbon, and oxygen and finally, $p+N^{15}\rightarrow C^{12}+He^4$, predominates in larger stars. For these reactions, the parent nuclei must have sufficient activation energy to overcome repulsive electrostatic forces and to come into and to be confined in contact rather than being scattered, allowing them to be fused by the strong nuclear force. The required activation energy is temperature of the order of 10^7 K. At these temperatures, the reactants are plasma—ions and electrons fully dissociated in rapid random motion. Confinement in stars is by very high gravitational forces. Rate of reaction is relatively slow. So far, developed use of artificial fusion is limited to explosive rates of reaction in thermonuclear weapons. A U235 or PU239 fission trigger is arranged to heat and compress fission fuel, deuterium, tritium, or lithium, which transmutes to tritium under neutron bombardment; ${}_3Li^6 + n \rightarrow {}_2He^4 + {}_1T^3$, long enough for a fusion reaction to occur. Controlled fusion reactions for power production have huge potential as an inexhaustible energy source that does not produce long-lasting radioactive waste, but they are much more difficult and not yet achieved. The fusion process must be self-sustaining and produce more thermal energy than is lost, primarily by radiation. Stellar process cannot be reproduced and research is based on use of artificial reactions. Energy release is a maximum with the lightest elements and the highest rates of reaction are obtained with deuterium or tritium. The optimum reaction is $D^2 + {}_1T^3 \rightarrow {}_2He^4 + n + 17.6$ MeV. Even for this, temperatures $10x$ stellar values are required. In the absence of stellar gravity, only magnetic containment of low-density plasma under near vacuum is possible.

There are various schemes for a ring of plasma in a toroidal containment. The Tokamak, using a combination of toroidal fields produced by external electromagnets and the helical field component due to the plasma current, is the most promising. Heating to 10^{8}°C is ohmic, induced by a transformer arrangement in which the plasma ring is the secondary, supplemented by the injection of beams of high-energy neutral fuel atoms and by microwave heating. In the D–T reaction, approximately 80% of energy release is kinetic energy of neutrons, which are not affected by the containment. In a future power plant, this will be absorbed by a moderator in sections of a surrounding blanket and transferred to the working fluid of a power cycle. Other sections of the blanket will breed tritium fuel from lithium. Although radioactivity will be hundreds of times less than for a comparable fission reactor, intense neutron radiation will render structural materials radioactive and reduce their integrity. So far, the highest ratio of heat produced to heat input Q is 1.25. The highest Lawson number—the product of plasma density and containment time—achieved is 10^{13}. $Q = 10$ and a Lawson number of 10^{14} are required for a practical power plant.

NUCLEAR SAFETY

The nuclear industry grew very rapidly under strong development pressures and several accidents involving core meltdown and fission product release—the most serious of which are described in this article—occurred. Increasing concerns regarding all levels of radiation release have resulted in an industry culture giving safety overriding importance. The International Atomic Energy Agency provides overall surveillance, standards, information, and advice. National governmental bodies such as the U.S. NRC set and enforce standards and license and monitor individual plants, including all aspects of ageing throughout their operating lives. Licensing requires a rigorous proactive analysis process to identify all hardware and human incidents with potential for accident. All aspects of design, construction, and operation are of the highest possible standards to minimize their probability, but successive countermeasures are designed to prevent fission product release if they occur. The discipline, effort, and expertise devoted to safety are the highest in any industry. Protection against fission product release involves (i) containment within fuel rods and prevention of overheating. (ii) Redundant systems to ensure safe shutdown and decay heat removal. (iii) Primary containment to prevent a loss of coolant accident (LOCA). (iv) Secondary containment to retain coolant and emergency core cooling systems (ECCS) to prevent core meltdown in the event of a LOCA. (v) Secondary containment to contain a core meltdown. Primary containment is the reactor pressure vessel or the system of pressure tubes that retain the reactor coolant. Secondary containment is a further pressure boundary surrounding the reactor, also designed to withstand external hazards such as aircraft impacts. Accidental formation of a critical mass and an uncontrolled chain reaction during a meltdown is not possible in a thermal reactor because of the dispersion of fissile materials within the core. Fast reactors have Core catchers designed to disperse and cool a meltdown. The technology of (iv) and (v) is complex, reliant on computer simulation, supported by simplified tests, and was the subject of intense, lengthy debate. The containment of meltdowns and fission products observed in accidents to reactors designed to NRC standards has reduced concern. The quest for increased safety continues. The 1 in 10,000 reactor year probability of core damage required by NRC is now reduced by a factor of at least 10 in current plants. The choice of characteristics that shutdown the reactor in the event of overheating, irrespective of control system or operator action, is a main objective of the designs now being developed and described under nuclear power plants.

NUCLEAR ACCIDENTS

The worst United Kingdom accident occurred in 1957 to a graphite moderated, open-circuit, air-cooled, plutonium production reactor at Windscale in NW England. A procedure requiring controlled temperature cycling to relieved energy stored in the graphite was incorrectly applied, initiating a fuel rod fire. This was revealed by radioactivity in the cooling air exhausted to the environment after only 40 hr, by which time 11 tons of uranium and much of the graphite was burning. Twenty-four hours of improvised firefighting using carbon dioxide and water were necessary to control it. Fission product release was limited by filters in the exhaust system but still contaminated some hundreds of KM^2. The major precautions taken were to prevent human ingestion of iodine131. Rigorous analysis was not done, but consequences appear to have been minimal. Both such reactors were decommissioned as a result. Neither the procedure that initiated the fire nor the open-circuit cooling is relevant to any present reactor.

In 1979, feed pumps in the secondary circuit of a PWR at Three Mile Island failed, preventing heat removal from the reactor cooling circuit via the steam generators. The reactor was shut down, but pressure increased due to decay heat release. A relief valve operated but jammed open, depressurizing the primary circuit. Instrumentation showed it to be closed. No core water level indication was available and indications from other parts of the primary circuit were false because of steam voids. Operators considered that levels were satisfactory and reduced water flow through the core when, in fact, a LOCA was occurring. Continuing steam evolution made primary circuit pumps ineffective and stopped all cooling. The core was exposed after 2 hr and about half of it melted. Radioactive steam and hydrogen were released into the secondary containment through auxiliary systems. After 7 hr, cooling water supply was re-established, enabling primary pumps to restart after 16 hr and the accident to be controlled. Some radioactive gases were vented in error to the environment. Consequences to the public were small. Maximum radiation doses were estimated as 0.1 rem, one-third of annual background levels. The reactor was irreparable and finally dismantled in 1993. This was the most significant nuclear accident in the United States and it had profound effects. Damage to public confidence (later exacerbated by Chernobyl) effectively terminated new plant construction in the United States for two decades. Operator training, safety procedures, and redundancy of control and instrumentation of existing plant were urgently improved. The incident demonstrated the vulnerability of light water reactors to a LOCA. Even after shutdown, decay heat overheats fuel in 1–2 min compared to 10 times longer in a gas-cooled reactor. However, primary and secondary containment remained intact, a partial melt-down was contained, and radioactivity release resulted only from human error, increasing confidence in the ability of current reactor design to confine the consequences of accidents to the plant.

Ironically, an experiment intended to increase safety caused the world's most serious reactor accident. The RMBK natural uranium-fuelled graphite-moderated pressure tube BWR at Chernobyl in the Ukraine was one of four at the site and of 14 such Russian designs built in the countries of the former Soviet Union. The experiment was intended to show that if the turbine generator were tripped, in the absence of grid supplies, its stored rotational energy was sufficient to power reactor cooling pumps until emergency diesel generators could be started to do so. During preparation, some safety and control systems, which would have interfered with the test, were disconnected. Because of this, power level could not be reduced accurately and fell below the 30% intended. It could not be restored because of xenon poisoning and additional neutron absorption due to an increase in coolant, as the pumps to be used in the test were started. In the attempt to do so, more control rods were withdrawn than permitted by operating procedures. To start the test, grid driven cooling pumps were shut down and steam flow to the turbine stopped. The latter and the resultant run down of the pumps reduced reactor cooling, increasing steam in the core. The RMBK design has a positive void coefficient—reactor power increases as steam content in the coolant increases. This is opposed by the fuel Doppler effect—reactivity reduces as fuel temperature increases. The net effect, power coefficient, depends on the power level. At high load it is negative, but at the 20% load at which the test was conducted it is positive. Thus, reactor power increased rapidly, causing operators to attempt to SCRAM the reactor. Control rods had graphite ends. As they were reinserted from their overwithdrawn position, their initial effect was to displace coolant (a neutron absorber) with graphite (a moderator), accelerating instead of reducing the power increase. Severe heating then distorted channels, jamming the rods and prevented the SCRAM. Power increased exponentially, melting fuel and causing explosive steam evolution, which burst the lightly constructed reactor casing and building. Ingress of air initiated a graphite fire, which was the major cause of release of most of the fuel and waste products into the environment.

Station staff, firefighters, and cleanup crew acted with extraordinary courage and selflessness, most receiving lethal radiation doses. There were 56 immediate deaths. The final consequential toll is uncertain, but estimates vary from 9000 to 90,000. Although this accident is probably the most important determinant of world public opinion, it is significant only in emphasizing the overriding responsibility to ensure safety. The experiment was poorly planned, the safety case for it was inadequate, and it proceeded although the planned conditions were not achieved. The power characteristics of the reactor were dangerous and unknown to operating staff. Adequate secondary containment was not provided. In all respects, operation and design grossly violated the safety culture discussed above.

EFFECTS OF RADIATION

All radioactivity is damaging to living tissue. Alpha particles are relatively massive and cause severe damage to tissue they encounter, but rapidly loose their energy in doing so. They are stopped by skin and they are dangerous only when alpha emitters are absorbed by inhalation or ingestion, and in some cases, subsequently concentrated within the body. Beta and gamma radiation are ionizing—they produce ions of the substances they encounter by removing electrons, both harmful to tissue. Beta is more damaging but is stopped by thin metal. Gamma radiation is extremely penetrating and requires shielding by materials of high atomic numbers, such as lead, to reduce intensity to acceptable levels. Because it has no charge, neutron radiation is not directly ionizing, but its absorption makes a nucleus emit ionizing radiation. For the same reason, it is also extremely penetrating and best shielded by a mass of light nuclei. In practice, several feet of concrete are required.

In the United States, the unit of absorbed radiation energy is the Rad (radiation absorbed dose, equal to 0.01 J/Kg). To account for different damage potentials of types of radiation, exposure of living tissue is measured in rem (roentgen equivalent man) = absorbed radiation (rads) multiplied by a factor. This is 1 for X, Gamma and Beta, 20 for Alpha, and 5–20 for neutron radiation, depending on its energy. SI units are the gray (100 rad) for absorbed radiation and sievert (100 rems) for equivalent dose. Knowledge of the effects of radiation is from nuclear weapon and accidents victims, radiographers, and nuclear industry workers. Effects on individuals vary widely, but the following is indicative: acute doses over 300 rem cause skin burns, immediate damage to blood, and digestive and central nervous system failure, and are fatal within weeks. Twenty rem received over a short period significantly increases the risk of leukemia and other cancers. Much lower levels can cause genetic damage, possibly mutation. Time between exposure and its effects depends on dose rate but there are no certain thresholds below which no harm will occur. Unavoidable public background exposure depends on location. In the United States, annual values are maximum 1 and the national average is 0.37 rem, composed approximately as shown below. An intercontinental flight can result in a dose of 0.004 rem.

Natural radioactivity				Man-made radioactivity	
Radon gas from decay of uranium	**Cosmic rays**	**Terrestrial radioactivity**	**Radioactivity in the body**	**Medical x-rays**	**Other**
50%	10%	10%	10%	10%	10%

The U.S. NRC has set annual limits over background radiation of 0.1 rem for the public and 5 rem for workers who work with radioactive materials. Limits remain controversial because of the difficulty of isolating and confirming the long-term effects of low levels of radiation.

CONCLUSION

Nuclear power offers nonpolluting energy supply in exchange for high standards of engineering, responsibility, and diligence. France, the United Kingdom, and the United States have operated large reactor inventories for over three decades with excellent safety records. The safety in depth philosophy, which is applied in the West, limits the increment of human radiation exposure due to nuclear power to an order less than natural levels and has prevented major fission product release in those accidents which have occurred. Chernobyl is irrelevant to any debate on the future of nuclear power other than as a demonstration of the consequences of repeated gross violation of that philosophy. This lesson is well taken worldwide, and the resurgence of nuclear power generation now beginning is well supported by innovative, safety-oriented nuclear technology.

BIBLIOGRAPHY

1. Jungk, R. *Brighter than a Thousand Suns*; Harcourt Inc.: San Diego, U.S.A., 1986.
2. Crease, R.; Mann, C. *The Second Creation; Makers of the Revolution in Twentieth Century Physics*; Rutgers University Press: New Jersey, U.S.A., 1986.
3. Lawrence, W. *Men and Atoms, The Discovery, The Uses and Future of Atomic Energy*; Simon & Schuster: New York, U.S.A., 1959.
4. Glasstone, S. *Source Book on Atomic Energy*, 3rd Ed.; Krieger Publishing Company: Florida, U.S.A., 1979.
5. Knief, A. *Nuclear Energy Technology, Theory and Practice of Commercial Nuclear Power*; McGraw Hill: New York, U.S.A., 1981.
6. Harvey, B. *Introduction to Nuclear Physics and Chemistry*; Prentice Hall: New Jersey, NJ, 1969.
7. El Wakil, M.M. *Principles of Nuclear Energy: Power plant Technology,* 1st Ed., McGraw Hill: New York, U.S.A., 1984; (Chapter 9), 354–394.
8. Pershagen, B.; Bowen, M. *Light Water Reactor Safety*; Pergamon: London, U.K., 1989.
9. Collier, J.G.; Davies, M. Chernobyl. In *Greenay*; Harrison Ltd.: Gloucester, U.K., 1986.

Nat–Perf

Performance Contracting

Shirley J. Hansen
Hansen Associates, Inc., Gig Harbor, Washington, U.S.A.

Abstract
Energy efficiency (EE) reduces operating costs and frees up funds for capital improvements. It also makes funds available to purchase text books, buy new medical equipment, or hire more teachers. It lowers costs for consumers, enables enhanced industrial competition, and has the potential of significantly reducing energy-related pollution.

With all of the incredible benefits of EE, the question inevitably arises: why isn't EE being aggressively pursued around the world? In survey after survey, when owners and operators are asked why they do not perform EE work, the biggest reason is *money*—or the lack thereof. Even if an organization has the requisite money to do EE work, there is always competition for the available dollars in a budget.

A little analysis reveals that we usually lose the budget fight because top management seldom views EE as an investment, not an expense. Typically, it's viewed as an expense compared to other organizational expenses. Since EE measures are seldom viewed as central to an organization's mission, our EE budget arguments suffer.

Overcoming such a major obstacle requires considerable finesse. It helps to step back and look at the situation from management's point of view. To reach them, it makes sense to position EE as a way to reduce operating costs through self-funded projects. Even if we get management thinking along these lines, however, the initial investment is still a problem. Unfortunately, this is frequently compounded by a lack of technical expertise, which, in turn, adds to management's discomfort. This discomfort is usually manifested in a query similar to, "How do you know it will save enough to pay for the project?"

A wonderful answer to all of these concerns is embedded in an industry that has the necessary expertise, can deliver the initial capital to make EE happen, and furthermore guarantees it will work. This industry offers management a clean deal of "reduced operating costs without capex." (Capex is management shorthand for capital expenditures.) But the words from this industry that really make management smile are, "We don't get paid unless we deliver what we promised." The concept is known as performance contracting, defined as a contract with payments based on performance and managed by an energy service company (ESCO).

INTRODUCTION

Historically, performance contracting has been based on guaranteed future energy savings, although it is increasingly applied to water, waste water, O&M, and other applications. Performance contracting allows the customer, e.g., an industry, state agency, hospital, school or commercial business, to use future energy savings to upgrade facilities and cut operating costs.

An ESCO, which provides the performance guarantee, will inspect a facility and/or an industrial process for energy-saving opportunities, recommend EE measures, and implement those measures acceptable to the owner at no upfront capital cost to the owner. The ESCO then guarantees that the value of the energy savings will cover the cost of the capital modifications and services if the cost of energy does not fall below a specified level.

Performance contracting is not new. The idea was born over 100 years ago in France, with a focus on district heating efficiencies. Royal Dutch Shell saw the concept's potential and exported it to the United Kingdom and the United States. Today, the performance contracting concept has gained traction around the world.

Initially, Royal Dutch Shell, through its subsidiary, Scallop Thermal, offered to deliver all needed energy services to a customer for 90% of its current utility bill. By delivering such services for a cost below the 90% figure, Scallop Thermal recovered its operating costs and made a profit. From just such an effort at Hanneman Hospital in Philadelphia, the idea of shared savings was born in the United States. These early projects were based upon each party sharing a percentage of the energy cost savings generated by retrofits and referred to as "shared savings". During the life of the contract, the ESCO expected its percentage of the cost savings to cover all of the costs it had incurred, plus yield a profit. This concept worked quite well, as long as the energy prices stayed the same or escalated.

But in the mid-1980s, energy prices dropped. Suddenly, it took longer than expected for an ESCO to recover its costs. With markedly lower energy prices, paybacks often became longer than the contracts. Some firms could not

Keywords: Performance contracting; Shared savings; Guaranteed savings; Energy Service Companies; ESCOs; Energy efficiency financing.

meet their payments to suppliers or financial backers. Several ESCOs closed their doors and, in the process, defaulted on commitments to their shared savings partners. "Shared savings" was in trouble—and the process became tainted by lawsuits and suppliers' efforts to recoup some of their losses, while facility managers tried to explain their own losses, which were previously guaranteed.

To make matters worse, it was discovered during this troubling time that one of the ESCO pioneers, Time Energy, had been entering into shared savings with an eye toward benefiting primarily from federal investment tax credits and energy tax credits. The building owner did not necessarily receive any energy cost savings benefits. Furthermore, the head of Time Energy got in trouble with the security and exchange commission (SEC). Stories traveled and soon the trust vital to accepting a new concept was badly shaken.

Fortunately, many ESCOs persisted in their efforts to make the new concept work. Some projects continued to show savings benefits to both parties. Of even greater importance, several companies, which had guaranteed savings, made good on those guarantees.

In spite of this tenuous start, the "shared savings" industry survived, but its character changed dramatically. Those supplying the financial backing and/or equipment recognized the risk of basing contracts on energy prices. With uncertainty in the industry and greater uncertainty in energy pricing, risk levels grew and interest rates went up. By 1990, the use of true shared savings agreements had shrunk to approximately five percent of the U.S. market.

In its place, new names, new terms, new types of agreements, and different financing mechanisms emerged. In part to respond to the negativity that surrounded the term, "shared savings," the industry focus turned to guaranteed performance. And *performance contracting* emerged as the favored name, along with a new model called "guaranteed savings."

PERFORMANCE CONTRACTING TODAY

From its shaky beginnings to its near death when oil prices plunged in 1986, a strong performance contracting industry has emerged. Performance contracting is gaining acceptance around the world and is expected to continue to change and evolve.

ESCOs today offer a broad range of retail energy services, including:

- Engineering feasibility studies
- Equipment acquisition and installation
- Load management
- Energy supply; power marketing
- Facilities management and water management
- Outsourcing
- Risk management
- Indoor air quality services
- Measurement and savings verification
- Energy information management
- Environmental compliance
- Guaranteed results

With regard to measurement and verification (M&V), it should be noted that some ESCOs provide this service, but customers should be aware that an inherent conflict of interest exists when the party receiving the payment also determines the extent of that payment. There are third-party firms in the United States that provide M&V service, such as MMSI in New Jersey and the Texas A&M Energy Center. In any case, it is strongly recommended that the M&V work comply with internationally accepted guidelines, such as the International Performance Measurement and Verification Protocol (IPMVP). (IPMVP guidance documents can be downloaded from the organization's web site: ipmvp.org).[1]

Following the patterns we have all witnessed in the telecommunications industry, ESCOs began—and continue—to unbundle and bundle their services, offering several or all of the services listed above to their customers. Ultimately, ESCOs are apt to sell conditioned floor space, as ESCOs and end users look for more efficient means of guaranteeing a return on their investment. Such an advent will bring us back full circle to the first supply/demand efficiencies agreement at Hanneman Hospital.

We can now expect to see the concept of energy performance contracting extended to other areas, such as resource management. As previously noted, several ESCOs now offer water management on a performance contracting basis. Just as performance contracting has already been extended to encompass operational savings (first in the United Kingdom), performance contracting will work in any setting where one can define the parameters, establish a baseline, and deliver the service cost-effectively for less than that baseline.

In the meantime, those now considering performance contracting—as a consumer, an ESCO, a financier, or utility—have a history rich with experience. We now have the ability to look at fully implemented projects, which have put the theory of performance contracting into practice.

HOW ESCOs WORK

Through the years, ESCOs have come to recognize that performance contracting is primarily risk management. Much of their work and the structuring of the deals, therefore, focus on effective risk management and mitigation.

The first critical step, then, is the qualification of customers who pose risks for the ESCO. Attractive customers (Energy management and ESCO selection from the customer's perspective are presented in *Manual for Intelligent Energy Services*,[2] which is fully referenced at the end of this section) must offer more than savings opportunities; they must have the qualities needed for a successful long-term partnership. The criteria, which include organizational characteristics, such as business longevity, and facility factors, such as projected use patterns, vary slightly among ESCOs.

But the process of selection is viewed by all successful performance contractors as a vital first step.

Structuring the Deal

Once the customer has been selected and the general concept is agreed upon, the deal is typically structured as follows:

- A Planning Agreement is signed.
- An investment grade energy audit (An investment grade audit is critical to ESCO and project success. A book on the topic is referenced at the end of this section) is performed.[3]
- Owner approves measures to be implemented.
- An energy services agreement (ESA) is signed.
- Project implementation begins.
- The ESCO's project manager oversees the project for life of contract.
- O&M measures are performed as stipulated in the ESA to assure savings.
- M&V is performed; payments are made to financier.
- Annual reconciliation is completed to determine whether savings have covered debt service obligations. If not, the ESCO cuts a check for the shortfall.

Financial Models

Financing of performance contracts ranges from a manufacturer financing a piece of equipment on a "paid from savings" concept to an integrated business solution.[4] Each progressive step up the value chain, noted in Fig. 1, increases the complexity of the offering. At the same time, it also offers greater value to the customer.

One word in the above figure may need defining: "chauffage." This word is associated with the roots of performance contracting in France. It refers to heating in French. It is used here, however, as a label for a model wherein the ESCO provides the owner conditioned space for a fee per square foot (or square meter). The ESCO secures the energy supplies and makes sure they are used efficiently.

There are two dominant performance contracting models in the world. The first model was developed in France and is generally credited to Compagnie Generale de Chauff. With some minor modifications, this model is what we today call shared savings. For about ten years,

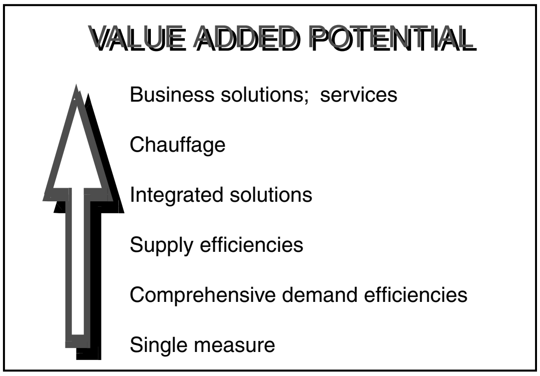


Fig. 1 The value-added chain.

shared savings was the only type of performance contracting offered in North America. The primary characteristics of shared savings are:

- The customer and the ESCO share a predetermined percentage split of the energy cost savings
- ESCOs typically carry the financing; i.e., credit risk
- Financing for the customer is often off balance sheet
- Equipment, which is often leased, is "owned" by an ESCO for the duration of the contract
- ESCOs carry both the performance risk and the credit risk
- Increased risks cause the cost of money to be higher
- Unless special safeguards are implemented, customers have greater payment exposure if energy prices or savings increase

The economic viability of shared savings rests on the price of energy. As long as energy prices stay the same or go up, the program can pay for itself.

After energy prices dropped in the mid-1980s, it was clear that the industry needed a model that no longer relied so completely on the price of energy to establish the project's economic viability. ESCOs in North America shifted to guaranteeing the amount of energy that would be saved, and further guaranteed that the value of that energy would be sufficient to meet the customer's debt service obligations as long as the price of energy did not fall below a stipulated floor price.

The significant characteristics of guaranteed savings are:

- The amount of energy saved is guaranteed;
- The value of energy saved is guaranteed to meet debt service obligations down to a floor price;
- The owners carry the credit risk;
- The risks to owners and ESCOs are less than with shared savings; and
- Less of the investment package goes to "buy" money.

Nat–Perf

While shared savings remains the dominant model in Europe, in the United States, roughly 90 percent of performance contracts are currently structured for guaranteed savings, with the owner typically accepting the debt through third party financing (TPF).

The typical cash flow of these two financing models is shown in the Fig. 2. In analyzing this cash flow, there are two distinguishing characteristics that should be noted.

First, in guaranteed savings, the ESCO and the lender seldom have a legal relationship. Usually, an informal relationship is established and certain conditions are understood. These conditions usually involve:

- Customer pre-qualification criteria
- Project parameters
- Stream-lined lending procedures
- Special interest rates

The second distinguishing characteristic appears in shared savings. In this case, the customer has no relationship with the lending institution and has little or no interest in the note being paid. Since all of the savings must happen in the customer's facility and/or process, this factor further raises the risks to the ESCO and the financier.

Another project financing model is emerging wherein the lender and ESCO create a Single Purpose Entity (SPE), which carries the credit and, to some extent, keeps the credit off the ESCO's books. In all cases, a parallel agreement is then secured with the ESCO to audit the facilities, install and maintain the equipment, provide other services, and guarantee that savings will cover required payments. This financial arrangement with a third party financier can be used as an approach with private sector customers, as well.

Conditions do exist that may temporarily encourage the shared savings model. In transitional economies, it is frequently difficult for customers to satisfy the bank's criteria for creditworthiness. As a new concept, performance contracting is easier to establish if the customer does not have to incur debt. Shared savings, however, relies heavily on ESCO borrowing and this presents a serious difficulty for small ESCOs, which lack financial resources. This further hampers industry growth. SPE allows an ESCO to satisfy these needs and to continue to complete projects, while avoiding becoming too highly leveraged.

Comparing the North American and French Models

Fig. 3, provides a side-by-side comparison of the two models. A unique characteristic noted in the graphic and often found in the French model is the four-step approach. This is worth calling attention to for the French four-step has two significant advantages: (1) it is truly self-funding; and (2) the opportunity to know and work with the partner before a lot has been invested reduces risks significantly.

This approach, often referred to in the 1970s as "roll-over" financing, typically uses the following pattern: Step 1, energy efficient operations and maintenance (O&M) measures are implemented; Step 2, savings from O&M measures fund quick fix (low investment) items; Step 3, savings from O&M and quick fix measures fund mid-range items (moderate capital costs and paybacks); and Step 4, savings from the first three steps fund the "big ticket" items (costlier measures and/or longer paybacks). As noted above, in the four-step method, risk management benefits are desirable and the projects generally do not require any outside financing. Under this approach, however, the end user must wait longer for major pieces of equipment and potential savings are lost in the interim. For the ESCO, the savings stream is slower to materialize.

Moving Up the Value Chain

Most financial models in use today are typically applied to comprehensive demand side management—the second level up the value chain. The next-highest level in the

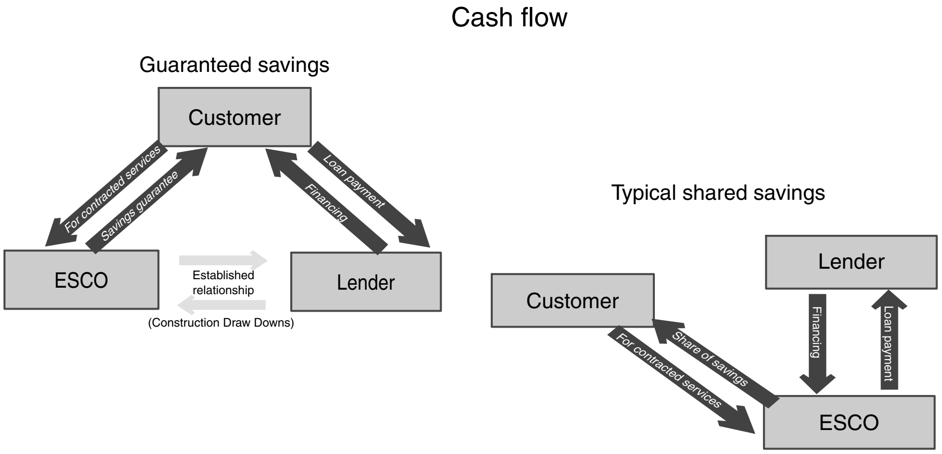


Fig. 2 Guaranteed and shared savings cash flow.

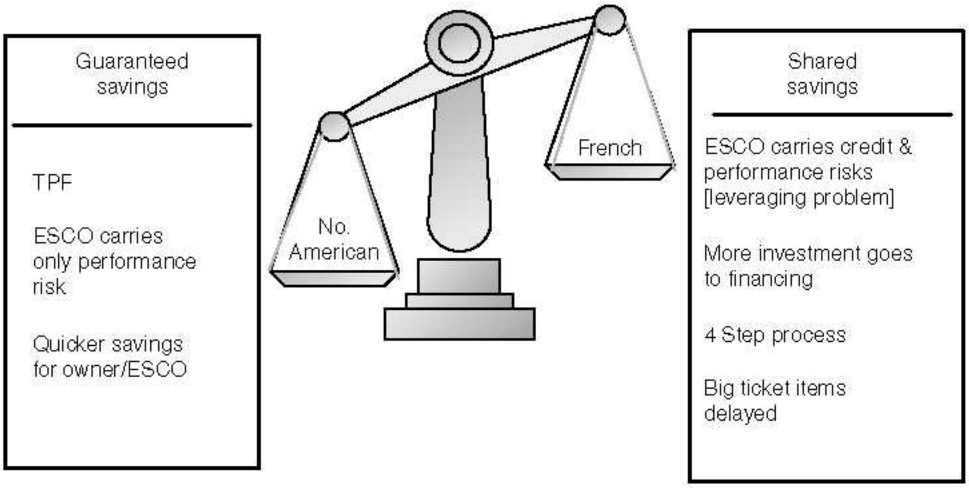


Fig. 3 Comparison of guaranteed and shared savings elements.

value chain is supply efficiencies. It is placed above comprehensive EE services only because the dollar amounts can be greater for work on the supply side of the meter.

When comprehensive EE services are paired with supply efficiencies, such as cogeneration or distributed generation, the package is referred to as an integrated solution.

Integrated solutions and chauffage are sometimes used interchangeably, but chauffage generally refers to a greater value-added approach. As mentioned earlier, under a chauffage arrangement, the ESCO conditioned space is provided for so much per square foot (or square meter). In such a case, the ESCO manages all supply and demand efficiencies. It may include some type of ownership of the supply generation or the HVAC system by the ESCO.

The ultimate value-added on the supply chain is the business solutions approach. Typically, this approach allows an ESCO to propose solutions that make prudent business sense, which may go beyond reduced energy consumption. The ESCO may provide services beyond energy efficiencies, wherein the energy cost savings may help defray the costs of this additional service. In other instances, the work may actually increase energy costs, but lower energy costs per unit of product through process efficiencies. For example, changing carpet curing procedures from a two-pass natural gas-fired oven to a single-pass electric infrared, as done at a North Carolina plant, increased total energy costs, but doubled production; so, the energy cost per unit of product dropped significantly.

PUSHING THE ENVELOPE

The "bottom-line" should clearly encourage the commercial world to pursue EE, whether the business is a manufacturer, industrial complex, commercial office, or institutional property. In his book *Lean and Clean Management*, Joseph J. Romm states: "Energy savings are bottom-line savings: depending on profit margins, that $5000 […in energy savings] could be better than a $100,000 increase in monthly sales, which entails increased costs for materials, labor, production, and overhead."

At the local government level in the United States, there are more than 4,000 performance contracts in place to assist public school systems in controlling energy costs and reducing energy consumption. Many cities and municipalities are also taking advantage of performance contracting benefits.

At the state government level, states continue to enact new, or improved, legislation, which encourages savings in public facilities and identifies performance contracting as an acceptable method of financing such projects.

On the federal government level, since the Energy Policy Act that passed in the early 1990s amended the earlier National Energy Conservation Policy Act, government agencies have been required to reduce energy consumption by a specified amount. These legislative actions have been paired with Presidential Orders that set specific targets for federal agencies to reduce energy consumption and have encouraged the use of performance contracting. Through the subsequent years, other statutes and orders have followed. Readers interested in the current status of such actions in the federal government are encouraged to contact the Federal Energy Management Program office in the U.S. Department of Energy for such information.

Internationally, the U.S. Agency for International Development has been particularly active and effective in making transitional economies aware of the advantages an ESCO industry can bring in the transition to a market economy.

The multilateral development banks, such as the European Bank for Reconstruction and Development, the World Bank, the Inter American Development Bank, and the Asian Development Bank, have all taken steps to foster and encourage the use of ESCOs in developing countries.

CONCLUSION

Performance contracting guarantees savings, and produces results. It recovers the money now going toward wasted energy and directs it to the organization's needs. It makes money while improving the environment, a simple concept. Unfortunately, it is a complex process. Too often, "Wishcos" attempt to do performance contracting without doing their homework. It is a growing, changing industry. As it evolves, ESCOs will broaden their offerings and bring new services to the deal. The global potential is huge.

If both parties take the time to understand the options and procedures, negotiate a fair contract, and exercise the necessary commitment, it will work.

And that's guaranteed.

REFERENCES

1. Energy Valuation Organization (EVO). International Performance Measurement and Verification Protocol. The document can be downloaded free from the web site, ipmvp.org. EVO contact can be made at P.O. Box 23363 Washington, DC, 2002; 6.
2. Hansen, S.J. *Manual for Intelligent Energy Services*, Published by the Fairmont Press. ISBN 0-88173-380-6 Distributed by Marcel Dekker: New York, 2002.
3. Hansen, S.J.; Brown, J.W. *Investment Grade Energy Audits: Making Smart Energy Choices*, Published by the Fairmont Press. ISBN 0-88173-362-8 Distributed by Marcel Dekker, Inc.: New York, 2004.
4. Hansen, S.J. *Performance Contracting: Expanding Horizons*, ISBN 0-88173-530-2 by The Fairmont Press in, 2006; 2nd Ed., Distributed by Taylor & Francis Ltd.

Performance Indicators: Industrial Energy

Harvey E. Diamond
Energy Management International, Conroe, Texas, U.S.A.
Robert G. Botelho
Energy Management International, Seekonk, Massachussetts, U.S.A.

Abstract
Industrial Energy Key Performance Indicators will address demand side IEKPI's—what they are, what types are utilized, the characteristics of their application, how they are determined, how they can be developed, what is needed to produce them, and how they can be used to bring energy savings.

INTRODUCTION AND SCOPE

As defined by www.management.about.com, "Key performance indicators (KPIs) are quantifiable measurements, agreed to beforehand, that reflect the critical success factors of an organization. Once an organization has analyzed its mission, identified all its stakeholders, and defined its goals, it needs a way to measure progress toward those goals. Whatever KPIs are selected, they must reflect the organization's goals, they must be key to its success, and they must be quantifiable."[1]

As defined above, KPIs are quantifiable measurements or metrics that are used to determine where important factors stand and what needs to be improved in order to achieve necessary goals when managing any activity. Key performance indicators are a key part of the effective management of any activity, whether it is a business or any other endeavor. It is not possible to effectively manage any activity without knowing where it stands or how well it is doing. It is not possible to effectively manage anything without monitoring it.

The importance factor, the critical success factor, or the main area of concern of an organization must first be clearly recognized and understood and then the specific goals related to improvements for the critical success factor must be identified in order to select and develop effective KPIs. In other words, effective KPIs must address a particular importance factor, a critical success factor, or an area of concern and then they must be specifically selected and developed to achieve specific goals relative to the critical factor or area of concern.

Industrial energy key performance indicators (IEKPIs) have two additional adjectives that specify that they apply to industry and more specifically to the consumption of energy in industrial manufacturing facilities. A few IEKPIs can be applied on the supply side of energy to evaluate short-term and long-term energy costs and contract issues, but by far, the majority of IEKPIs pertain to the demand side energy consumption, and this article will address only the demand side of energy consumption. The previously mentioned two adjectives indicate the critical success factor and the general goal for demand side IEKPIs—the amount of energy consumed in an industrial facility is the critical success factor for IEKPIs and the goal is to achieve the best possible efficiency and reduce any waste of energy for that facility. Industrial energy key performance indicators are quantifiable measurements within industrial manufacturing facilities that provide indications of how effectively and how efficiently an industrial manufacturing facility is consuming energy to manufacture products. The consumption of energy in industry today is definitely a critical success factor, with the cost of energy having risen to its present level. This article will address demand side IEKPIs—what they are, what types are utilized, the characteristics of their application, how they are determined, how they can be developed, what is needed to produce them, and how they can be used to bring energy savings.

Keywords: Industrial energy; Key performance indicators; Energy management; Efficiencies; Consumption analyses; Energy compilations; Energy balances; Process analyses; Key performance measurements; Mathematical models; Empirical determinations.

IEKPI DISTINGUISHMENT

Key performance indicators are being used to manage and increase the quality of production activities throughout industry today. Within an industrial facility, a KPI can be anything from the number of specification rejects per one million acceptable products, to the amount of feed material per pound of product, to the amount of man-hours applied per pound of product, to the amount of wash water per pound of product, to the amount of nitrogen utilized per pound of product, etc.

As mentioned earlier, IEKPIs in this article are aimed at the demand side of energy consumption in industrial facilities. Industrial energy key performance indicators are determined and quantified in order to effectively manage the energy consumption of an industrial facility. Industrial energy key performance indicators reveal and illustrate how effective an industrial facility is at using energy to manufacture products. Industrial energy key performance indicators reveal whether a facility is operating efficiently or if it is wasting energy. Industrial energy key performance indicators are focused on energy consumption by processes, systems, equipment, utilities, and all other aspects of an industrial manufacturing facility. Industrial energy key performance indicators are focused on energy consumption relative to production activities, but will also provide a measure of how well an entire industrial facility is using energy including its utilities, environmental, and administration areas.

Demand side IEKPIs should always address the amount of energy consumed and the efficiencies for each item of equipment, system, unit, area, process, or any division of a plant that can indicate the need for capital projects or procedures that may be required to improve efficiencies and energy consumption, but they can also address the routine ongoing operation and maintenance of the plant.

In industrial manufacturing facilities, on the demand side, there are two major applications of IEKPIs—(1) that of identifying and making significant discrete changes to the design of the facilities and also to the operational or maintenance procedures that are being practiced within the facilities or (2) that of monitoring and managing the ongoing activities within the facilities. Industrial energy key performance indicators can reveal ways to save energy by making discrete changes within facilities but will also provide a basis for ongoing management in order to maintain consistency in effective use of energy.

There are many different types of industries that utilize different forms of energy and different processes in order to manufacture many different types of products. However, there are many commonalities between industries in the types of processes, types of facilities, types of equipment, and the operational and maintenance procedures that are applied. There are common energy issues within industries and there are common ways to apply IEKPIs to save energy.

When an IEKPI has been determined and quantified within an industrial facility, it should be compared to a standard or a benchmark value in order to indicate and identify possible improvements or managerial directions that need to be taken. Depending on the specific IEKPI, the benchmark value might be that of an industry, process, facility, or equipment energy consumption standard. Sometimes when there is no industry, process, or equipment energy consumption standard available it is necessary to determine the standard by performing energy engineering calculations based on the operating variables involved and the application of realistic factors that can determine a realistic level of energy consumption for the object of measurement. Or, if sufficient historical data can be collected, an empirical determination can be performed. Industrial energy key performance indicators are always compared to a standard or benchmark in order to provide valuable changes or managerial instructions for the improvement of energy consumption for the object of measurement.

TYPES OF IEKPIs

The types of demand side IEKPIs that are applied in industry today are: efficiencies, energy consumption analyses, energy requirement compilations, energy balances, process analyses, key performance measurements (KPMs), mathematical models, and empirical determinations. A discussion will now follow that will outline what each type of IEKPI is and how it is applied in industry today.

Efficiencies are pretty well understood today as generally being the ratio of the amount of effective work produced by an object to the amount of energy consumed by the object in producing the work, whether it be a piece of equipment, a facility, a process, or a plant in general. The predominate application of efficiencies is toward evaluating an object for discrete improvement, whether it be design changes or revised operating and maintenance procedures. An example would be evaluating the efficiency of a boiler and determining that revisions need to be made to its construction or determining that the boiler needs to be operated or maintained differently. Efficiency calculations for boilers, combustion systems, steam systems, lighting, and other systems can be found in the *Energy Management Handbook* by Wayne C. Turner and Steve Doty[2] and in the *Guide to Energy Management* by Barney L. Capehart, Wayne C. Turner, and William J. Kennedy.[3] The evaluation of the efficiency as an IEKPI would entail referring to a benchmark standard for efficiencies of the same type boiler that is used widely throughout industry. Efficiencies can also be utilized in monitoring equipment or facilities on an ongoing basis in order to continuously manage the operation and maintenance of them, but usually it requires distinct testing and additional effort in order to measure efficiencies and efficiencies are normally not measured as frequently as would be required for ongoing management. Efficiencies are more often utilized for discrete changes as described above, and ongoing management considerations are usually covered by KPMs that will be discussed later.

Energy consumption analyses utilize engineering calculations in order to determine how much energy is being consumed by a piece of equipment, a facility, a processing unit, or an area of a plant. Many example engineering calculations of energy consumption for many forms of energy for various types of thermodynamic and fluid-flow equipment are provided in *Fundamentals of Classical Thermodynamics* by Gordon J. Van Wylen and Richard E. Sonntag.[4] Electrical engineering energy consumption calculations for electrical equipment can be found in *Basic Electrical Engineering* by A.E. Fitzgerald, David E. Higginbotham, and Arvin Grabel.[5] These analyses can be performed in two different ways and for two different purposes. Energy consumption analyses are frequently performed in order to determine the benchmark standard itself by calculating how much energy the object should be consuming using the proper values of the process variables involved and by applying realistic factors to determine a realistic energy consumption level. An example of this would be calculating the amount of heat that would be realistically required (not just theoretically, but including a realistic factor) in a process dryer in order to dry away so many pounds of water that is contained in the product that is being dried as it enters the dryer. Once this calculation has been made, the calculated results can then be compared to the actual amount of heat that is being used to dry the same amount of product. If there is a wide difference between the two numbers, this is an indication of potential energy savings and further investigation could possibly identify projects and procedures that can be implemented to reduce the energy consumption of this dryer to a realistic level. This application of an IEKPI would have provided its key results in energy management. Energy consumption analyses are also performed in a different manner whereby the actual amount of energy that is being consumed is calculated from just the values of process variables that are involved. This type of calculation is utilized to determine present energy consumption for an object when metering is not available for the object being studied. When the purpose is to determine the present level of energy consumption, there are two different applications that the results can be used for—to determine if discrete changes or improvements are needed and for the ongoing management of energy for the object. When the application is for the purpose of evaluating the object for discrete improvement, the results of these analyses will be compared to an industry standard or facility benchmark for evaluation. When the application is for the purpose of ongoing management, the results of these analyses will be compared to the historical value benchmark in order to identify management directions.

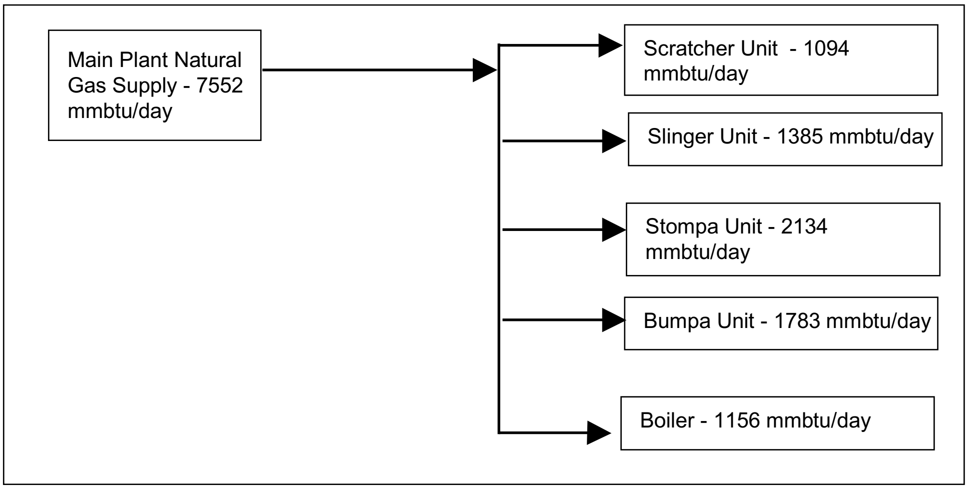
Energy requirement compilations are performed on a facility to determine how much energy the facility should be using based upon the equipment that uses energy and the applicable usage factors and load factors. A good example is trying to determine how much electricity a factory should be using based upon how many electric motors are installed, how many motors are used during normal operation, and how much load each motor should be pulling when normally running for production within the facility (see Table 1 below for an example). An energy requirement compilation involves listing all of the energy-consuming equipment, applying the realistic and applicable usage factors for each piece of equipment, and applying the proper load factor for each item of equipment. Another example of a possible energy requirement compilation would be that of the steam heat exchangers in a facility that determine how much steam is normally used in the facility. Normally energy requirement compilations are performed in order to determine the benchmark standard of energy consumption for a facility. This benchmark standard can have application for determining discrete improvements of making capital project changes to the facility or making changes to the operating and maintenance procedures. This benchmark standard can also be used for frequent comparisons with actual usages under an ongoing energy management program to identify proper management directions on an ongoing basis.

Energy balances determine how much energy is being consumed in each part or division of a facility. Energy balances simply provide energy consumption amounts for a division of a plant based upon the amounts that are being consumed in other parts of the plant (see Table 2 below for an example balance). Energy balances can sometime provide IEKPIs in a facility at a higher level of examination than most of the other IEKPIs, but most usually they are applied in order to determine the benchmark level of energy consumption for an area or

Table 1 Total plant electrical motor compilation

Motor no.	Description	H.P./kW	Load factor	Daily usage factor (h)	Total daily (kWh)
1A	Water supply pump	100/74.6	0.8	20	1193.6
3G	Tank mixer	30/22.38	0.85	24	456.552
5B	Process feed pump	200/149.2	0.75	12	1342.8
7C	Reactor agitator	20/14.92	0.85	18	228.276
9K	Product pump	150/111.9	0.7	9	704.97
Total plant daily kWh					3926.198

Table 2 Plant natural gas balance



unit within a facility. When used as an IEKPI, energy balances can indicate whether or not an area or unit of a plant is using more energy than it should, but at this level of examination, it may not be as effective in indicating improvements or management directions. Nonetheless, it is possible that an energy balance can indicate wastage of energy if it is compared to an energy requirement compilation and indicates a significant difference in consumption. Usually when energy balances provide an indication of possible improvement, further investigation and additional IEKPIs are applied in order to determine and develop more specific improvements and directions.

One very important strategic indication that can be revealed by energy balances within an industrial facility is whether or not excess energy generated from the manufacturing processes in the facility remains in excess and that no effective energy savings can be produced by reducing the usage of this excess energy. An example of this would be a facility that recovers waste heat from its main processes in order to generate steam for other plant uses. The balance in this case indicates that there is more steam produced from the waste heat than is needed in the plant and excess steam is being blown-off or vented. This plant is producing more energy from its basic processes in the form of steam than it can use, and there is no way that energy savings can be accomplished by reducing steam usage in this plant.

Process analyses are performed on manufacturing processes in order to determine if a possible variation of the process could render greater energy efficiency without compromising product quantity, product quality, product specification, or safety, health, or environmental requirements. A process analysis could reveal that a certain process is not being operated under the most efficient operating conditions when considering all of the applicable process variables such as temperature, pressure, flow rate, liquid level, feed rates, catalyst amounts, process times, or any other possible process variables. Process analyses are most usually performed when the actual process conditions are already known, being the actual IEKPIs, and the process analysis will actually determine the benchmark or possibly a new standard that can be considered. This will most usually generate discrete improvements in either the facilities themselves or in the procedures of operation and maintenance. An example application of process analyses would be to conduct process modeling of a crude distillation tower in a refinery using PROMAX by Bryan Engineering[6] in order to determine if the best reflux ratio and reboiler rate is being used for the crude distillation tower.

When a process has been properly analyzed, its critical process variables have been identified with respect to their effect on energy consumption for the process, and the benchmark limiting values have been determined, the actual ongoing values of these process variables can possibly be applied as IEKPIs in order to properly control the process for energy consumption. An example of this would be when it is determined that a process temperature should be realistically within a certain range considering all process requirements, this temperature is monitored to ensure that it remains within the proper range, and immediate corrective actions are taken whenever this temperature varies outside the acceptable limits. This would be an ongoing application IEKPI in order to maintain energy conservation.

Key performance measurements are the most effective IEKPIs for generating ongoing energy management directions for continued energy conservation within industrial facilities. Key performance measurements are ratios of energy consumption to production quantities for a plant and any of its divisions—facility, unit, area, or individual major equipment. However, the more drilled down they are applied (like down to major equipment items) the more effective they are in providing insight into the waste of energy and possible improvements. Key performance measurements require actual, accurate, and concurrent measurements of energy consumption and production for any object that is to be studied and managed. Metering of energy consumption and frequent totalization of consumption are required for the energy consumption data so that concurrent data can be placed into the KPM ratio of energy consumption to production for any object that is being monitored. When KPMs have been quantified, they are normally compared to historical records for comparison to a historical benchmark level but they can also be compared with a strategic industrial or facility benchmark in order to determine if management directions are needed and should be implemented immediately.

If KPMs are applied at a higher level, like for an entire plant or facility, they may not be valuable for indicating useful numbers for analysis of energy consumption ratios at such levels because of the possible presence of base load amounts in the energy consumption numbers. If a significant base load of energy consumption (energy that is consumed by the facility whether or not the facility is under production) is present within a facility and is not accounted for, it can produce indications that are not accurate for the variations in energy consumption relative to variations in production levels. For example, if a facility had a base load of energy consumption of thirty percent of the normal consumption for the facility and then the production and the actual variable energy consumption both increased by thirty percent, the indication of the Gross KPM would be erroneous. It would indicate that the energy consumption per unit of production had gone down by almost seven percent and that the facility is producing more efficiently (see Table 3 below for illustrative calculations). Whenever this factor is present and is of significant concern a Net KPM should be utilized where the base load of energy consumption is accounted for (here is an application where an energy balance and an energy requirement compilation could be helpful in arriving at useful numbers) and removed from the Gross KPM numbers. Going back to the previous example, the Net KPM would indicate no variation in ratio of energy to production, which would be accurate if no other variables were involved.

Another problem with Gross KPM numbers when applied at higher levels in attempting to analyze trends in energy consumption per unit of production is that other variables can have a significant influence while not being accounted for, which can produce false analysis for the Gross KPM. Frequently, ambient conditions and possibly other numerous factors play a significant role in determining energy consumption in a facility, depending upon the characteristics and nature of the facility. If these variables or drivers are not accounted for, the resulting Gross KPMs that could be used for analysis can be misleading and produce false indications. When this situation exists and an accurate analysis of energy consumption is required, a different IEKPI is needed—a mathematical model of energy consumption that provides a benchmark of energy consumption according to production levels, ambient conditions, and all other specific variables applicable to the facility being analyzed—sometimes called a "PREUFTOOL."[7] This mathematical model has to be developed from an engineering analysis of the energy consumption in a facility and conformed to actual data for the energy consumption, production, ambient conditions, and all of the applicable variables or drivers that are involved for the facility. Once this mathematical model has been developed, it can then be used as a benchmark for energy consumption for a facility and it will calculate how much energy should be consumed by the facility according to its past characteristics and according to the value of all of the influencing variables.

Sometimes when not analyzing at such a high level but more so at the equipment or process facility level, empirical determinations of energy consumption characteristics can be useful for determining benchmarks. Empirical determination of energy consumption within an industrial facility can take place by collecting KPMs for the facility over a period of months and years so that the variations of the empirical values can be tracked and analyzed and conclusions can be determined. Empirical determinations of energy consumption will usually vary over time due to the effect of numerous variables that are not constant over the time interval. However, by tracking the empirical determination over a significant amount of time, it is possible to observe correlations with variables

Table 3 Illustraction of inaccuracy of gross macro key performance measurements (KPMs)

Facility has a 30% base load of energy consumption—30% of energy is always being consumed, irregardless of production levels or even if there is no production
Production increases by 30% and is now at a level of 130% of the previous level
The actual variable level of energy—70% of normal energy consumption when there is production—increases by 30% (70% × 1.3 = 91%)
The total energy consumed by the facility is now 91% + 30% base load = 121% of previous level of energy consumption
The Gross Macro KPM of amount of energy per unit of production is now 121% of energy divided by 130% of production = 0.930769 and is down by approximately 7% from the previous level. This is indicating that the facility is operating more efficiently when the facility is operating at the same ratio of actual variable energy to production

and to arrive at values concurrent with operating conditions. An example of this would be to collect energy consumption and steam production for a boiler over a period of months for both winter and summer conditions. By correlating the amount of Btu's per pound of steam to the applicable steam generation rates and the ambient temperature, it can be possible to determine the amount of Btu's per pound of steam for various steam generation rates and at various ambient temperatures if no other significant factors, such as variable feed water temperatures, come into effect during this time. Once this graph or table of energy to steam ratios has been determined, it is possible to determine the amount of energy that the boiler should be consuming for a certain steam generation rate. This same concept could be applied to the production of products in a production unit of a manufacturing facility to determine the benchmark amount of energy per unit of production in correlation with the many variables that might effect energy consumption over a period of time. Conducting this determination of an empirical benchmark may indicate that possibilities exist for improved energy consumption and that further investigation could identify projects or procedures that can reduce energy consumption for the facility. This application of a KPM, IEKPI would have provided its key results in energy management.

DEVELOPMENT OF IEKPIs

Development of IEKPIs should be based upon knowledge of the facility that is to be studied and managed. Knowledge of industrial facilities comes by virtue of study of process flow diagrams, process analyses, piping and instrument diagrams, electrical one-line diagrams, control system diagrams, equipment drawings and specifications, standard operating procedures, maintenance records, on-site familiarity, and on-site observations of each of these sources of information. Knowledge of the facility should be acquired first and then the main factors that determine and define IEKPIs for a facility can be rightly developed. Industrial energy key performance indicators should be site-specific and should be tailored to the characteristics of the facility that is to be measured and analyzed for energy consumption performance.

Industrial energy key performance indicators that address discrete changes to the facilities and to procedures and that address ongoing energy management should be developed and applied for each area or unit of a manufacturing facility and should be applied down to the equipment level as much as possible. In general, the further down into a facility that IEKPIs are developed the greater the effectiveness will be realized in improving energy consumption and eliminating waste. For instance, KPMs can be very useful when applied down to the equipment level. As an example, the amount of Btu's per pound of steam produced is a very revealing IEKPI for boilers and should be monitored frequently on an ongoing basis for the operation of a boiler.

Industrial energy key performance indicators that identity discrete changes that may be needed for the facility in the form of capital projects should always be examined early in an overall energy conservation program within an industrial facility because these items usually identify large levels of energy savings and it takes longer to get these projects completed to realize the savings than other efforts that can be implemented in a shorter time frame. The application of energy balances, efficiency determinations, energy consumption analyses, and process analyses should usually be conducted first, respectively in the order mentioned, so that any needed major revisions are recognized first within a facility.

ITEMS NEEDED FOR IEKPIs

Items that are commonly and frequently needed for IEKPIs are: submetering for units, areas, or major energy-consuming equipment; accurate and concurrent consumption data for divisions of a facility that are to be monitored; historical process data from a process historian that has been recording process variables over time intervals; process flow diagrams and process design criteria for all processes; equipment specifications and characteristics; and energy engineering equations and calculations that can be used to calculate energy consumption of processes and equipment.

Automatic data links for metering data and for process variables to be downloaded from a process historian will greatly aid the production of IEKPIs and will enhance their accuracy and consistency for controlling and managing energy consumption in an industrial facility. Automatic data systems that produce IEKPIs will consistently run smoother and be more easily applied in managing energy consumption in an industrial facility.

HOW TO USE IEKPIs

After IEKPIs have been produced they should be reviewed by management for indications of measures that can be taken to increase efficiencies and reduce waste of energy. As mentioned earlier in this article, there are two major applications—(1) discrete major changes to the facilities or to procedures for operating and maintaining the facilities and (2) ongoing management of operations and maintenance of the facility. The first application includes two different types of measures—(1) discrete changes that are usually capital projects that will make energy consumption improvements in the facilities and (2) revised operating and maintenance procedures that will improve energy consumption on a day-to-day basis and eliminate waste. The IEKPIs that are more applicable to the first application are: energy balances, efficiencies, energy consumption analyses, and process analyses. These IEKPIs will usually indicate what factors need to be improved in a process, facility, or item of equipment and will usually also indicate what actual changes can be considered. The second major application of IEKPIs is actually the largest recognized application of KPIs in industrial energy consumption. Key performance measurements are the IEKPIs that are utilized for this major application—that of managing ongoing, day-to-day operations and maintenance. The reason that this is actually the largest application of IEKPIs is that there are so many KPMs that can be defined and utilized in an industrial facility for each unit, area, or major item of equipment. Multiple KPMs can be utilized for just one major item of equipment, for an area of production, or for a processing unit.

When KPMs have been produced and placed into an easily reviewed graph, they should be reviewed by management for the identification of any inconsistencies in energy consumption. If an inconsistency is discovered, the nature of the operation and the maintenance of the area, unit, or major equipment should be reviewed to determine if there is a correlative and identifiable activity on the part of operations or maintenance that could be the cause of the inconsistency. Correlation of activities to discovered inconsistencies has been very effective in identifying causes of energy waste within industrial facilities and has enabled managements to correct operational and maintenance activities. This application and effective use of KPMs as effective IEKPIs has been very successful in maintaining energy conservation within industrial facilities and has actually identified additional energy saving measures in some instances. Again, KPMs are probably the major application by numbers of IEKPIs within industry.

NEED FOR ENERGY MANAGEMENT IN INDUSTRY

The cost of energy has risen so much in the past few years and continues to rise even higher at the present. The result is that the cost of energy consumption in industrial manufacturing facilities has become a critical success factor for many businesses and may be a predominate factor in some industries in determining whether or not the business is even profitable or not. This importance, as well as the additional importance of environmental effects such as reducing green house gases, and the importance of keeping businesses viable so that citizens have job opportunities would seem to make prominently known the need for effective energy management within industry today. However, it seems that the presence of corporate politics within businesses today can frequently maintain a desire to keep things going according to the status quo, so that someone's present agenda will be more likely to bring them recognition and success in their own mind. Corporate politics seems to frequently favor not examining the actual status of things today, but to keep things reputable according to their present thinking. The hope is that self-gratifying corporate political thinking will give way to positive thinking that will bring possibilities for improvement. It is very possible that all of industry today could reduce their energy consumption by five to ten percent by applying IEKPIs for the identification of discrete changes to their facilities and to operational and maintenance procedures and for an ongoing management of energy consumption through the application of KPMs.

SUMMARY

Industrial energy key performance indicators can be used to manage energy consumption within industrial manufacturing facilities for the purpose of identifying discrete major changes to the facilities themselves and to operational and maintenance procedures but also by the application of KPMs in managing ongoing day-to-day operation and maintenance of industrial manufacturing facilities. The application of all types of IEKPIs can bring significant energy savings to industry today and should be considered by all industries for specific application within each facility. Hopefully, all of industry will use IEKPIs to find ways to improve facilities and to manage energy consumption in order to make their businesses more viable and to improve the environment of the world in which we live.

REFERENCES

1. Reh, F.J. Guide to Management, Key Performance Indicators, http://management.about.com/cs/generalmanagement/a/keyperfindic.htm
2. Turner, W.C.; Doty, S. *Energy Management Handbook*; Fairmont Press: Lilburn, GA; 6th Ed.; ISBN: 0-88173-542-6, 2005.
3. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. Guide to Energy Management; Fairmont Press: Lilburn, GA; 5th Ed.; ISBN: 0-88173-477-2, 2004.
4. Van Wylen, G.J.; Sonntag, R.E. *Fundamentals of Classical Thermodynamic*; Wiley: New York, 1967.
5. Fitzgerald, A.E.; Higginbotham, D.E.; Grabel, A. *Basic Electrical Engineering*; McGraw-Hill Book Co.: New York, 1967.
6. Bryan Engineering. PROMAX; BRE: Bryan, TX, 2005; http://www.bre.com
7. Energy Management International, Inc. PREUFTOOL, http://www.wesaveenergy.com

Encyclopedia of

Energy Engineering and Technology

First Edition

Volume 3

Pages 1147 through 1667
Phot–Wire

Photovoltaic Systems

Michael Ropp
South Dakota State University, Brookings, South Dakota, U.S.A.

Abstract
Photovoltaics (PV) is the direct conversion of sunlight to electricity. This entry contains a brief review of the history of PV; a survey of solar resource determination options; a summary of first, second and third-generation PV devices and concentrating PV devices; overviews of the power electronics (charge controllers and inverters) and batteries associated with PV systems; a discussion of PV systems and their design; and a brief examination of PV applications, with comments on future prospects.

HISTORY

Photovoltaics (PV) ("photo" meaning "light" and "voltaic" referring to electricity) is the direct conversion of sunlight into an electrical potential (a photovoltage) that can be used to provide electric power. The PV effect itself was discovered in 1839 by a French physicist, Edmund Becquerel, who observed that a photocurrent would flow between two electrodes in a solution when the apparatus was exposed to light.[1] Later, the effect was noticed in selenium by William Adams and Richard Day,[2] and the first solid-state solar cells were made from selenium by Charles Fritts and Werner Siemens.[3] However, many investigators were skeptical about these devices because the quantum physics required to explain the observed effect were not known yet. It wasn't until Max Planck's proposal of the quantum nature of light in 1900[4] that the theoretical foundations for understanding PV were established.

Selenium-based PV remained impractical because of its high cost and low efficiency (about 0.5%).[5] Then, in 1954, Calvin Fuller and Gerald Pearson were working on new silicon rectifier diode technology at Bell Laboratories, and during one experiment they found that their device produced a significant photocurrent when strongly illuminated.[5] At that same time, Daryl Chapin was working on selenium solar cells. When Pearson alerted Chapin to his silicon discovery, Chapin immediately abandoned his selenium work and switched to silicon, and after significant effort but a relatively short time, the result was the achievement of 6% conversion efficiency.[6] However, the energy cost for PV, which is the critical figure of merit for PV systems (usually expressed in $/kWh), was nearly a thousand times that of competing alternatives at that time.[7] Although technically successful, PV was still too expensive to be useful.

Fortunately, another PV application was identified—power for satellites.[8] Its reliability, relatively high specific power (power per unit weight), and lack of a need for refueling in orbit made PV a natural choice for spacecraft power. Vanguard I, in 1958, was the first PV-powered satellite.[9] The satellite power market supported the fledgling PV industry, which in turn aggressively worked to improve device performance. Efficiencies climbed to more than 10%[10] and energy costs declined. During the 1960s, a deep physical understanding of the operation of silicon solar cells was obtained by the PV research community,[11] a legacy that continues to pay dividends today. Still, PV's energy cost remained high, and terrestrial PV use was confined to small off-grid applications where other forms of energy supply or a grid connection were more costly than PV. Ironically, one of the biggest customers for the early terrestrial PV industry was the oil industry,[12] which used PV extensively in signaling systems for offshore oil drilling rigs.

Throughout the tumultuous 1970s and 1980s, the cost of PV-produced energy continued to fall, largely because of the growth of accumulated industrial and field experience, the "economies of scale" afforded by larger-scale production,[13] and the niche markets in which PV was cost-competitive continued to grow.

Then, in the 1990s, concerns mounted over the environmental impacts of the combustion of fossil fuels. This concern is perhaps best expressed by the term "the 3-E trilemma," explained by Yoshihiro Hamakawa as follows: we require the Energy needed to sustain Economic development without damaging the Environment.[14] An increasing awareness of the reality of the 3-E trilemma has spurred rising levels of interest in many forms of sustainable energy production, including PV. A variety of government programs worldwide, most notably in Germany and Japan, have spurred explosive growth in

Keywords: Photovoltaics; Solar electricity; Solar cells; Silicon; Batteries; Inverters; Charge controllers; System design.

the PV industry,[15] which in turn has led to higher levels of research support for PV and its associated technologies.

THE SOLAR RESOURCE

Photovoltaic systems use sunlight as their "fuel," and thus in PV system design it is crucial to know how much sunlight one might expect at the PV array site and how that sunlight is distributed day to day, season to season, and year to year. The terminology of solar energy can be a bit confusing because several commonly used but similar-sounding terms have critically different meanings.[16] The solar energy available on a site is quantified using the irradiation, which is an energy flux usually given in kilowatt hour per square meter per day and is also sometimes called insolation when referring specifically to sunlight. The solar power available on a site is given by irradiance, which is a power flux usually given in units of kilowatt hour per square meter. For PV system design, both of these quantities are important. The spectral properties of sunlight are also important to PV device operation, but at the system level this consideration is secondary.

The best way to gain an idea of what irradiation and irradiance can be expected on a site is to turn to a historical database of meteorological measurements taken at or near the site. Several such databases are available worldwide, and an ever-increasing number of solar databases are available online (although it should be noted that in many cases sunlight is not actually measured but rather modeled based on other meteorological measurements and observations). A few examples are:

- The Renewable Resource Data Center (http://rredc.nrel.gov), maintained by the National Renewable Energy Laboratory.
- Data from the Remote Automated Weather Station (RAWS) network are available from the Western Regional Climate Center (http://www.raws.dri.edu/index.html).
- The European Solar Radiation Atlas (see http://www.ensmp.fr/Fr/Services/PressesENSMP/Collections/ScTerEnv/Livres/atlas_tome1.htm).

If one requires solar resource estimates for sites far from any measuring station, there are procedures for computing such estimates based on a few simple inputs.[17,18] As one might expect, the accuracy of the computed estimates is not as good as estimates based on lengthy periods of historical data, but because of the high level of inaccuracy inherent in any long-term meteorological prediction, these computations coupled with some conservative design decisions are often quite sufficient for PV system design.

FIRST-GENERATION PV: CRYSTALLINE SI

Any PV cell has to accomplish two fundamental tasks: (1) it must absorb the energy of light and use it to generate mobile positive and negative electrical charges and (2) it must separate those mobile positive and negative electrical charges to produce a potential difference between them. There are many structures that can do this, with one of the most common being leaves—the process just described is a significant part of photosynthesis.[19]

Most man-made solar cells today are made from silicon (Si). Silicon is a semiconductor, which simply means that its electrical conductivity falls between that of metals and insulators. Silicon is in column IV of the periodic table, meaning that each silicon atom has four chemically-active electrons called valence electrons.

The detailed physics of solar cell operation are well described in many texts.[20–22] In the dark, each silicon atom's four valence electrons are tightly bound to their respective atoms so that they cannot move easily through the crystal. When photons of light strike the silicon, the photons' energy can be absorbed by the valence electrons, which are released from their bonds and become mobile charges. Also, each mobile electron leaves behind a "spot" where an electron could be, but isn't. These "spots" act for all practical purposes as if they were positive charge carriers or positive particles, and these positive charge carriers are called holes. Therefore, silicon can absorb light to create mobile electron-hole pairs (EHPs).

To separate the mobile charges, the silicon is used to make a structure called a diode. One key property of semiconductors is that their conductivities can be controlled through the addition of certain impurities, called dopants. In silicon, there are two main categories of dopants: elements from column V of the periodic table that have one extra electron and elements from column III that have one missing electron, or equivalently an extra hole. When added to the silicon crystal, the extra electron of the column V dopant becomes mobile in the crystal, and thus silicon doped with column V atoms will have an excess of mobile electrons. Similarly, the extra hole carried by the column III atoms is mobile in silicon, and a column III-doped silicon crystal will have an excess of mobile holes. We refer to the column V-doped silicon as *n*-type because its conductivity is controlled by the negative electrons, and column III-doped silicon is analogously called *p*-type.

The diode structure is shown in Fig. 1. It is essentially just a "sandwich" of an *n*-type and a *p*-type region. The interface between these two regions is called the *p*–*n* junction. Because of *p*–*n* junction physics, the junction acts as a one-way electrical valve; electrons can flow from the *p*-side to the *n*-side, but not the other way. Similarly, holes can flow from the *n*-side to the *p*-side but not back. Thus, when a photon is absorbed and generates an EHP, only one of the two charges can cross the junction, after which it can't get back. This separates the charges,

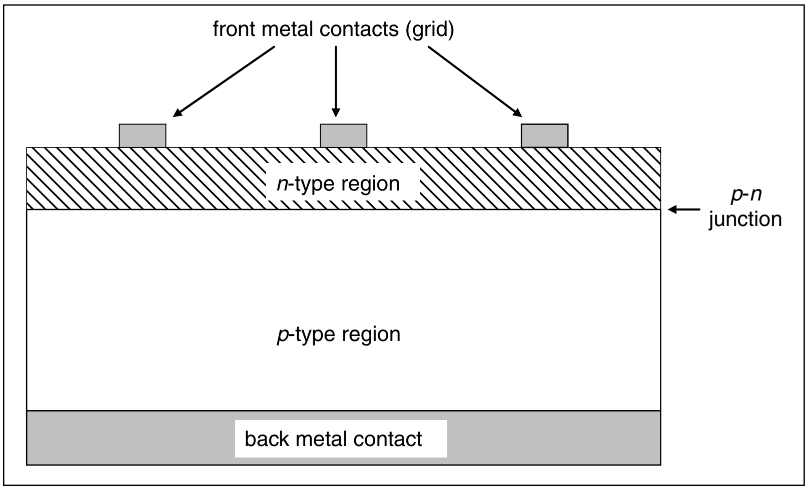


Fig. 1 Structure of a basic photovoltaic (PV) cell.

and a photovoltage builds up as one end of the device accumulates an excess of negative charge and the other end sees a buildup of positive charge. There is thus a positive and a negative terminal of the solar cell, just like a battery, and if electrical contact is made to the two ends of the device, the photovoltage will cause a current to flow. To turn the diode into a solar cell, adding electrical contacts is all that is required, remembering that the front contact must allow the light to enter the silicon. Usually a grid pattern is used for the front contact, as shown in Fig. 1.

Because this was the first "practical" PV technology, solar cells of the type described above are sometimes called first-generation PV devices. The structure in Fig. 1 is an extremely basic one. A wide variety of efficiency-enhancing features can be (and usually are) added to this structure.[23–25]

The efficiency of the above-described simple solar cell depends on many factors. One of the most important factors is the quality of the silicon material.[26] If the silicon is highly pure and has no defects in its crystal structure, then its electronic properties are favorable for high solar cell efficiency. Such material is referred to as single-crystal or crystalline silicon (c-Si). As might be expected, c-Si material leads to relatively high efficiencies but it is relatively expensive. Solar cells can also be made from lower-quality, less expensive materials such as multicrystalline silicon (mc-Si). In mc-Si, the silicon is composed of crystalline grains stuck together by regions of noncrystalline silicon. The grains have dimensions on the order of centimeters. Multicrystalline silicon is less costly than c-Si but has inferior electrical properties, and thus the efficiencies of solar cells made on mc-Si tend to be lower than those on c-Si.[27] These two types of silicon are shown in Fig. 2 below.

SECOND-GENERATION PV: THIN FILMS

First-generation PV devices suffer from a number of fundamental limitations. One of the most persistent is the aforementioned cost of the silicon material itself,[28] which combines with other factors to keep the energy cost of first-generation PV relatively high.

One fairly obvious way to reduce the cost of PV might be to reduce the amount of material required. Wafers of c-Si or mc-Si for PV cells are usually on the order of 150–300 μm (a micron is one one-millionth of a meter) in thickness. If it were possible to make solar cells much thinner but maintain high efficiency, then the material costs should be reduced.[29] This is the aim of so-called thin-film PV, or second-generation PV.

The basic structure of second-generation devices is usually essentially the same as that of first-generation devices, with a few modifications.[30] The main difference between first- and second-generation devices is that second-generation devices are not made from bulk crystalline material, but instead are fabricated on thin films of semiconductor formed using any of several deposition techniques, such as chemical vapor deposition (CVD).[31,32]

Silicon can be used to make thin-film devices. Using a variety of relatively inexpensive deposition techniques such as CVD, it is possible to create layers of silicon that are very thin—on the order of 10 μm—and make PV devices from these films. The resulting silicon usually has almost no crystal structure at all; it is neither crystalline nor multicrystalline, but amorphous silicon (a-Si). Amorphous silicon has very low electrical quality, which makes achieving high solar cell efficiencies difficult, although novel device structures and processing techniques have led to a-Si PV devices with efficiencies near 10%.[33] Unfortunately, a-Si-based devices have another drawback—they tend to degrade under illumination due to the Staebler–Wronski effect (S–W effect).[34] The physical mechanism behind the S–W effect is not yet fully understood, but it is well known that a-Si-based devices loses efficiency under sunlight exposure. Although thin film silicon devices still hold great promise because of their low-cost potential, considerable research

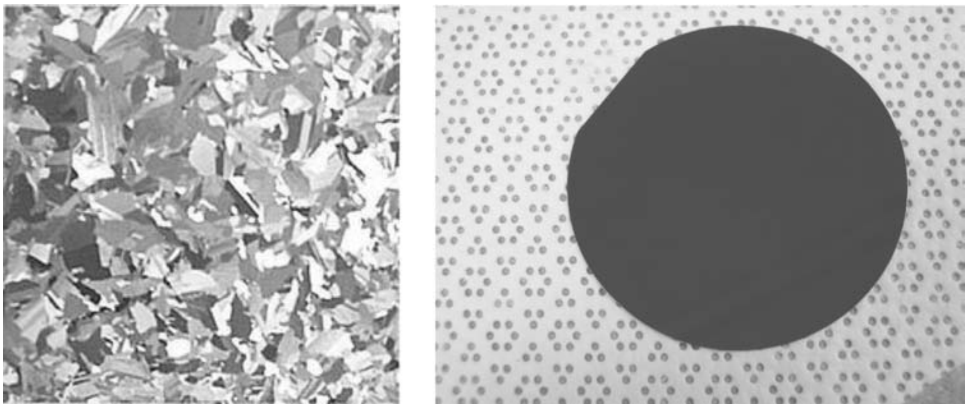


Fig. 2 Multicrystalline (left) and single-crystal (right) silicon wafers. Left photo courtesy of the National Renewable Energy Laboratory's Pictographic Information eXchange (NREL/PiX) (credit: Sandia National Laboratories); right photo by Jeff Moore.

remains to be done to improve their efficiencies and stability.

Silicon is not the only material that can be used to make second-generation PV devices. Successful PV devices have been made using a system containing cadmium telluride and cadmium sulfide layers (CdTe/CdS).[33] The chalcopyrite family is another family of alloys that has properties that could be very favorable for high-efficiency PV device fabrication.[35] This family includes copper indium diselenide (CIS) and a variety of quaternary alloys, such as copper indium gallium selenide (CIGS). Even in a-Si devices, some of the silicon layers are usually alloyed with germanium to produce SiGe.[36]

It is also possible to make second-generation PV devices using organic materials.[37] The advantages of using organic materials are that (i) the processing is relatively simple and inexpensive and (ii) the materials are cheap and readily available. Thus, in theory, these devices could have lower energy costs than PV cells based on inorganic semiconductors. That promise is yet to be realized, however, because the efficiencies of these devices remain low—in the range of 1%–5%.[38,39] Although organic devices must still produce mobile charge pairs and somehow separate the charges, the mechanisms by which they do this[37] are quite different than those of semiconductor-based devices, and are more akin to the process of photosynthesis.

THIRD-GENERATION PV: USING THE FULL SPECTRUM

Although there are a number of efficiency-limiting mechanisms in solar cells, both first- and second-generation PV devices suffer from two particular fundamental limitations.[40,41] First, there is a certain minimum photon energy, called the band gap energy and denoted E_g, below which the device cannot capture photons and convert them to electricity. Second, even if a photon does have energy greater than the band gap energy, E_g (actually, a bit less than E_g)[41] is the portion of the energy that the solar cell can convert to electrical energy. Those two loss mechanisms mean that first- or second-generation PV cells cannot use all of the energy of the solar spectrum.

It would be highly desirable to be able to circumvent this problem of inefficient use of the solar spectrum, and researchers worldwide are currently investigating approaches for doing that. The resulting full-spectrum PV devices are called third-generation PV devices.

One of the earliest and simplest approaches to improving the use of the solar spectrum is to use a multijunction PV cell or tandem PV cell[42] like that shown in Fig. 3. The multijunction device is basically a stack of cells designed such that each cell in the stack uses a different portion of the spectrum. Fig. 3 shows a three-cell stack. The materials of which the cells are made have different band gap energies, arranged from highest to lowest as shown. Cell 1, with the highest E_g (E_{g1}), captures the high-energy photons and can use most of their energy because of the cell's high E_g. The lower energy photons pass through it into Cell 2, which has an intermediate E_g (E_{g2}). Cell 2 absorbs midenergy photons. The photons with energy less than E_{g2} pass through Cell 2 and strike Cell 3, which has a low band gap energy (E_{g3}) and absorbs the low energy photons. The most efficient PV devices made to date (over 32%) are of this type,[43] and research is ongoing to continue to improve their performance and cost effectiveness. (It should be noted that almost all multijunction devices are also thin-film devices, which blurs the line between second- and third-generation PV.)

Other novel approaches are being enabled by advancements in our understanding of extremely small structures, with dimensions on the order of a few nanometers (so-called "nanostructures"). Nanostructures can behave very differently than larger-scale structures in some

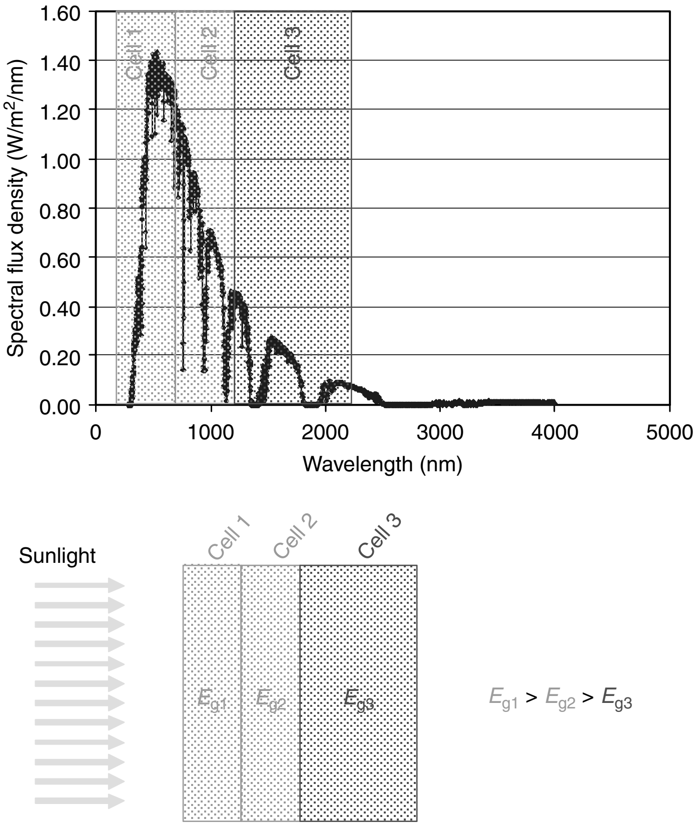


Fig. 3 Schematic representation of a multijunction photovoltaic (PV) cell.

respects, because at those small dimensions quantum physical effects begin to dominate the structure's behavior. One aspect of nanostructure behavior that is particularly important for PV is that the optical properties of nanostructures begin to change and become dependent on the size of the nanostructure. This potentially means that the optical properties can be controlled by simply controlling the size of the nanostructure. With such controllability, it may be possible to create devices that in effect split higher energy photons into multiple lower energy photons, or sum low energy photons into a single higher-energy photon. A device incorporating this feature would shift the energy of the solar spectrum into ranges that are used with maximum efficiency, thereby leading to significantly higher overall device efficiencies. Examples of such devices include:

- Intermediate band devices,[44,45]
- Hot-carrier cells,[46,47]
- Impact ionization devices,[48,49] or
- Devices utilizing spectral upconversion or spectral downconversion.[50]

Many researchers worldwide are pursuing third-generation devices, and initial results indicate that third-generation PV devices with efficiencies much higher than those of first-generation devices may be available in the future.

CONCENTRATION (FOCUSED SUNLIGHT)

One approach to reducing the cost of PV is to concentrate the sunlight. In this way, the total required area of expensive solar cells can be reduced, being replaced by large-area but less-expensive lenses or mirrors.[51] Also, most PV cells become more efficient under concentrated sunlight.[52] Concentrators have been shown to be capable of reducing PV energy costs.[53] Most concentrators require fairly accurate solar tracking, although one type, the luminescent concentrator, does not.[54] Generally, concentrators are feasible only in larger-scale PV power plants,[51] but significant research efforts are currently underway to continue driving down the energy cost from concentrating PV systems.

FROM PV CELLS TO PV POWER PLANTS

Consider the single silicon solar cell in Fig. 4A below. As a general rule of thumb, this cell typically produces on the order of 0.5 W at a voltage of about 0.5 V, depending on

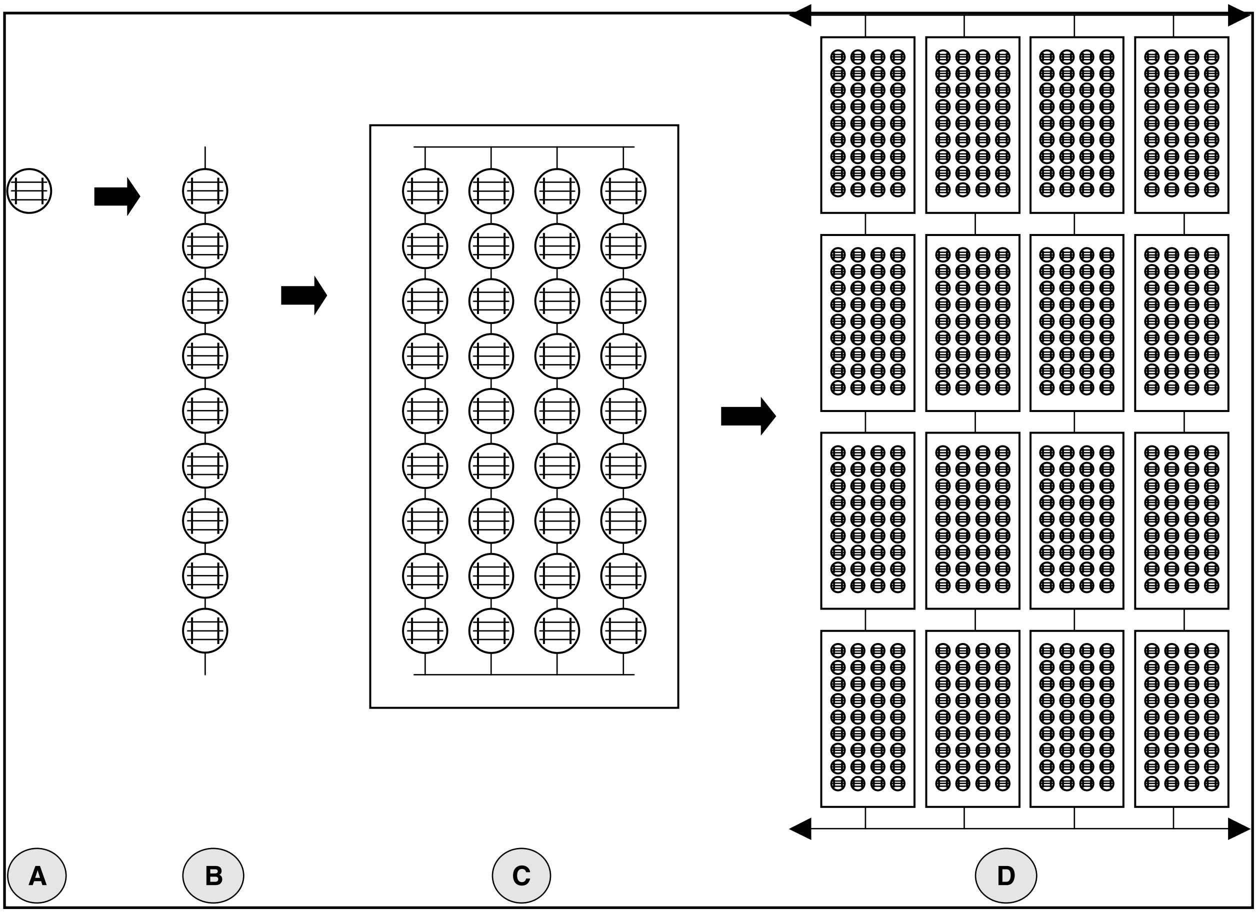


Fig. 4 Individual photovoltaic (PV) cells (A) are combined in series strings (B) to obtain higher voltage. Series strings are encapsulated together to form a module (C). Modules are connected in series strings, and module series strings are connected in parallel to form a PV array (D) that produces the desired voltage and power.

the amount of sunlight and the cell temperature. This is inadequate for most applications. To obtain higher voltages, solar cells are typically connected in series to form series strings, as shown in Fig. 4B. In silicon PV, series strings commonly contain around 36 cells in series.[55] Series strings can be connected in parallel to increase power output. A group of one or more series strings is then encapsulated in a module. Fig. 4C shows a schematic of a module and Fig. 5 is a photograph of a module at work in a water pumping application. The module typically has a glass front sheet, a plastic backsheet, an aluminum frame, and a terminal box that provides a means for electrically connecting to the PV cells. Modules are available in a wide array of sizes, from less than 5 W to over 300 W. Finally, to meet the voltage and power needs of a specific application, modules are connected in series and parallel as shown in Fig. 4D, forming a PV array.

The number of PV modules required in a series string or array depends on a number of factors, the most important of which is the energy requirement of the electrical load. Larger loads require more PV modules, but one must also remember that sunlight is a statistical parameter. Photovoltaic systems are usually designed to provide a specific level of availability (percent of time that the load is being powered—in other words, the percent of time during which there is not a power failure).[56,57] Cloudiness affects availability, as does the seasonal distribution of the load; if the load is larger in the wintertime when there is less sunlight, availability suffers. A high availability requirement will also increase the needed number of modules. Several references contain details on how to size a PV system for a given application.[58–62]

Photovoltaic arrays only provide electricity when the sun shines, so to provide for the needs of most loads a complete PV system must also contain either energy storage or backup generation. There are three basic PV system configurations, shown in Figs. 6–8 below. Fig. 6 shows a very general configuration of a stand-alone PV system in which 100% of the load energy must be supplied by the PV array, and the array must be designed according to load energy and availability requirements. The stand-alone system usually includes batteries to provide continuous power to the load, but not always; for example,

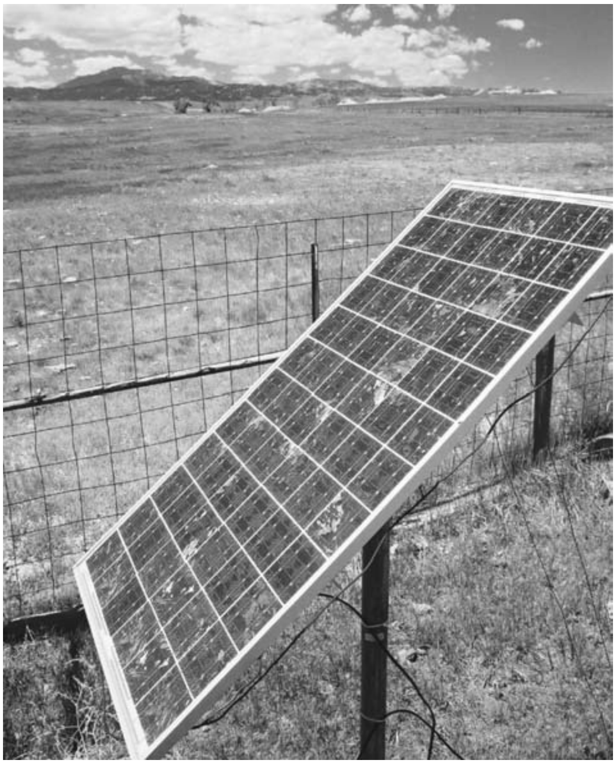


Fig. 5 Photovoltaic (PV) module in the field (this one is in a livestock watering application near Laramie Peak in Wyoming). Photograph by Susan Ropp.

water pumping applications, such as the one shown in Fig. 5, usually do not use batteries because in those cases the water can be stored.

Fig. 6 shows two different loads—one that requires DC power and another requiring AC power. The DC load often is powered directly from the battery bank, but if it requires a highly regulated voltage or a voltage different than the battery bank voltage there may be a DC–DC converter, as shown. Because PV arrays produce DC power, AC loads necessitate the inclusion of a DC–AC converter or inverter. In a system of this type, the inverter's output is AC voltage.

Stand-alone PV systems can be cost effective if the load is relatively small and if the load is sufficiently far from the grid that a grid extension, transformer, and service drop become prohibitively expensive. However, if either the load energy demand or required availability is too large, the large PV array needed may make the cost of a stand-alone system prohibitive. One solution to this problem is shown in Fig. 7, which shows a hybrid PV system configuration. The hybrid system includes an additional generator of some type, so that the PV array does not have to supply 100% of the load's energy needs. The system in Fig. 7 includes an engine-generator set (E-G). Gasoline, diesel or liquified propane gas (LPG)-powered generators are all used in hybrid PV systems, and in some cases wind turbines or small hydropower units can also be used. In the hybrid system, the PV array can be much smaller because it does not have to meet the entire load energy or availability requirement. Instead, the entire hybrid system can be subjected to an optimization process to determine the PV array size and generator run strategy that leads to the system with the lowest energy cost. This is a nontrivial process, and several software packages are available to assist with it.[62] Sometimes a simpler process can be used, in which the PV array is designed to minimize generator run time within some specified cost constraint.

In most locations in the industrialized world, the energy cost from PV is higher than that from the utility grid. However, this is not always the case. Some locations have exceptionally high grid-based energy costs; in others, special policies or incentives have been put into place to encourage the use of renewable energy. In these cases, one might find the system configuration shown in Fig. 8, a grid-connected or utility-interactive photovoltaic (UIPV) system. In a sense, the UIPV system is a special case of the hybrid system shown in Fig. 7. Utility-interactive photovoltaic systems use the utility grid as their backup generator and usually do not incorporate local energy storage (batteries). The lack of batteries greatly simplifies the system, as seen in Fig. 8. The point labeled "PCC" is the point of common coupling between the PV system and the grid.

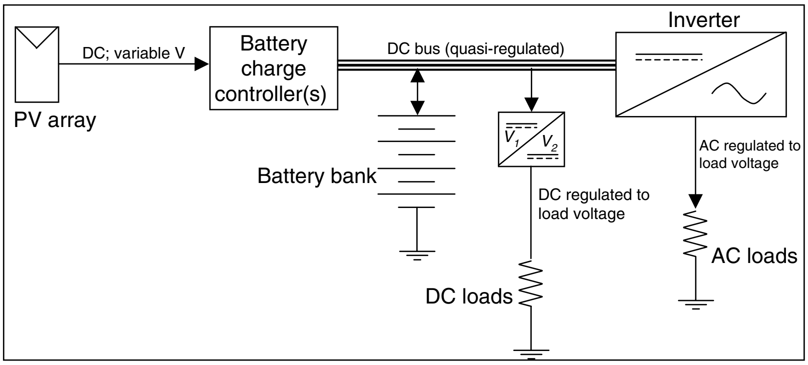


Fig. 6 Generalized configuration of a stand-alone photovoltaic (PV) system.

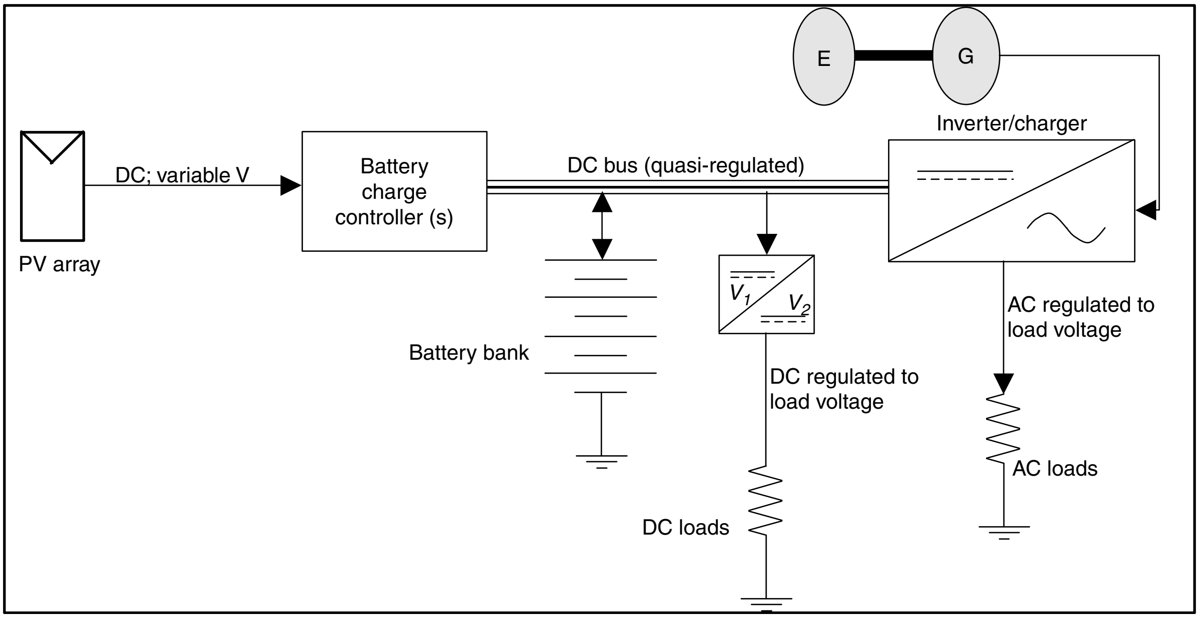


Fig. 7 Generalized configuration of a hybrid photovoltaic (PV) system, in this case using an engine-generator set (E-G) as the redundant generator.

PV POWER ELECTRONICS

As is clear in Figs. 6–8, power electronic converters, or so-called switch-mode power converters, are critically important in PV systems. The battery charge controllers and inverters fall into this category.[63–65]

Charge controllers are typically DC–DC converters (their inputs and outputs are both DC, but at different current/voltage levels—see Figs. 6 and 7). Inverters, by definition, have DC inputs but AC outputs (see Figs. 6–8). As shown in Fig. 7, some inverters are bidirectional in that power can flow from DC to AC (inverter mode) or from AC to DC. This latter mode is used to import power from the hybrid generator to recharge the batteries, and thus it is called the charger mode.

Photovoltaic cells have one operating point (that is, one voltage and current) at which they are maximally efficient. Thus, one job of PV power electronics is to attempt to keep the PV array operating at that highest-efficiency point, or maximum power point, at all times. Specialized control algorithms called maximum power point tracking (MPPT) algorithms accomplish this task. The MPPT algorithm would be in the charge controllers in Figs. 6 and 7 and in the inverter in Fig. 8.

The charge controller must also manage the charging and discharging of the batteries. As will be discussed below, batteries must be charged and discharged carefully to maximize their lifetimes and minimize operational problems. Charge controllers must balance the battery's needs with the requirement for MPPT. Notice that MPPT is not always possible—if the batteries are fully charged and the electrical load is being satisfied, then even if the PV array is capable of producing more energy, that extra energy has nowhere to go, and the PV array must be operated at a point other than its maximum power point.

The inverter's primary job is to supply AC power. In Figs. 6 and 7, the inverter produces an AC output voltage. The wave shape of the output voltage may be a sinewave, as in utility systems, but it may also be a square wave or a so-called modified sine wave (see Fig. 9). Square wave and modified sine wave inverters are generally more efficient

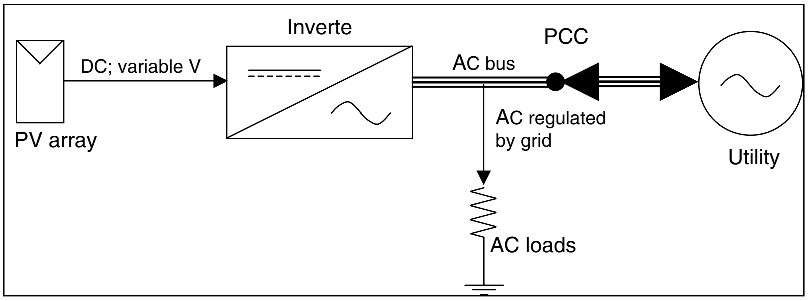


Fig. 8 Generalized configuration of a grid-connected photovoltaic (PV) system.

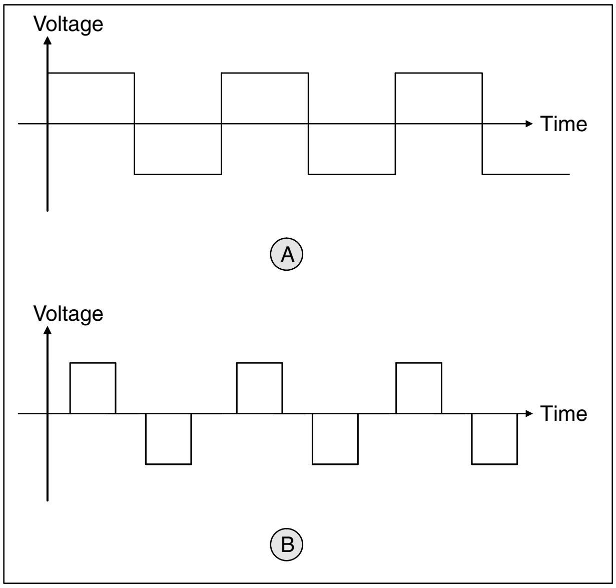


Fig. 9 Two common nonsinusoidal inverter output waveforms. (A) square wave. (B) so—called "modified sine wave."

and less expensive than sinewave inverters, and they are used whenever the load can tolerate the nonsinusoidal waveform. In Fig. 8, the inverter's output voltage is fixed by the utility grid, and thus the inverter's output in that case is an AC current (not voltage) and must be sinusoidal.

ENERGY STORAGE FOR PV

Photovoltaic arrays only convert energy when light strikes them, so it is necessary to use energy storage to provide power to loads at night and during extended cloudy periods. (Note that PV arrays do work when there are clouds; they simply don't produce as much energy on cloudy days.)

The most common type of storage in PV systems is lead-acid batteries.[66–68] There are several types of lead-acid battery used in PV. Flooded batteries use a liquid electrolyte. They are the longest lasting (if maintained properly), but water must be added to them periodically. It is possible to use special valves to reduce the need for addition of water, and batteries so equipped are called valve-regulated lead-acid (VRLA) batteries. These require less maintenance but are not as long-lived. One can also use different types of electrolyte, such as the gel used in gel-cell batteries. Gel cells require essentially no maintenance but are shorter-lived than VRLA.

Lead-acid battery technology is mature, but it is a bit challenging to use.[68,69] If they are discharged too deeply, they permanently lose storage capacity. If they are overcharged, some of the electrolyte can be vaporized and lost. The charging and discharging rates must also be carefully controlled to avoid permanent loss of capacity.[60] If any of these requirements are not met, the battery lifetime will be shortened.[68,69] Also, when they are discharged they can freeze at a relatively high temperature, so they must be protected against this.[69] Still, they are widely available and they are the least expensive type of battery,[68] so they continue to be the most common type used in PV.

Other types of batteries can be used. For example, pocket-plate nickel–cadmium batteries have some properties that are better than lead-acid for PV applications,[68,69] but Ni-Cads are becoming less available because of concerns over the toxicity of cadmium. Nickel metal hydride (NiMH) batteries retain some of the advantages of Ni-Cads without using cadmium, and thus NiMH may be a promising candidate in some applications. Both Ni-Cads and NiMH batteries are significantly more expensive than lead-acid at this time.

Eventually, hydrogen may prove to be the ultimate storage medium.[70] Photovoltaic could be used to electrolyze water into its hydrogen and oxygen constituents, which could be converted back to electricity by a fuel cell or combustion turbine or used to power vehicles. Although very promising because of its sustainability, its potential to displace oil use, and its reliance only on very

Fig. 10 A photovoltaic (PV)-powered rail yard switch near Alliance, Nebraska (Photo by Susan Ropp).

Fig. 11 A photovoltaic (PV)-powered instrumentation package near Gordon, Nebraska (Photo by Susan Ropp).

common resources, hydrogen storage is too expensive to be feasible at present.

PV APPLICATIONS

As mentioned before, the key figure of merit for PV energy is an economic one—the energy cost, usually in dollar per kilowatt hour. It is common practice to compute the life-cycle cost or levelized energy cost of the PV system, which is the energy cost considering all PV-related costs over the entire lifetime of the system.[71]

Even with current technology, there are many applications today in which PV is cost-competitive.[72–74] In particular, small off-grid loads are good candidates for PV. These include autonomous instrumentation systems, highway and marine navigation and signaling systems, remote communications repeaters, and livestock water pumping. A few examples are shown in Figs. 5, 10–13. In addition, in the developing world, almost two billion people live without electricity and are effectively “off grid.” Photovoltaic can be cost effectively used to provide them with basic lighting, refrigeration, water pumping, and communications.

However, the present energy cost of dispatchable PV (that is, PV + batteries) is still on the order of five times higher than that of electricity from the grid. This situation is changing, though; the cost of externalities and other

Fig. 12 Traffic signaling applications: a temporary stoplight in a construction zone in Brookings, South Dakota (Photo by Michael Ropp).

Fig. 13 Traffic signaling applications: a sign warning of an upcoming intersection near Punkin Center, Colorado (Photo by Susan Ropp).

costs will continue to add to the cost of fossil-fuel power while technology and economies of scale continue to drive PV costs down. The range of applications in which PV is cost-competitive continues to expand, and perhaps one day dispatchable PV will reach a cost at which it provides a significant portion of the energy we get from the grid. Energy storage remains a critical need, but new advancements in PV device technology give good grounds for optimism that at least that part of the system costs will someday reach these goals.

REFERENCES

1. Zweibel, K. *Harnessing Solar Power: The Photovoltaics Challenge*; Plenum Press: New York, 1990; 2.
2. Perlin, J. The Story of Solar Cells. In *Organic Photovoltaics: Mechanisms, Materials and Devices*; Sun, S.-S., Sariciftci, N., Eds., Taylor and Francis: Boca Raton, FL, 2005; 3.
3. Komp, R. *Practical Photovoltaics: Electricity from Solar Cells*; Aatec Publications: Ann Arbor, MI, 1995; xii–xiii.
4. Messenger, R.; Ventre, J. *Photovoltaic Systems Engineering*, 2nd Ed.; CRC Press: Boca Raton, FL, 2004; 15.
5. Perlin, J. The story of solar cells. In *Organic Photovoltaics: Mechanisms, Materials and Devices*; Sun, S.-S., Sariciftci, N., Eds., Taylor and Francis: Boca Raton, FL, 2005; 4.
6. Zweibel, K. *Harnessing Solar Power: The Photovoltaics Challenge*; Plenum Press: New York, 1990; 4.
7. Zweibel, K. *Harnessing Solar Power: The Photovoltaics Challenge*; Plenum Press: New York, 1990; 5.
8. Zweibel, K. *Harnessing Solar Power: The Photovoltaics Challenge*; Plenum Press: New York, 1990; 6.
9. Perlin, J. The story of solar cells. In *Organic Photovoltaics: Mechanisms, Materials and Devices*; Sun, S.-S., Sariciftci, N., Eds., Taylor and Francis: Boca Raton, FL, 2005; 7.
10. Green, M. *Silicon Solar Cells*; Centre for Photovoltaic Devices and Systems: Sydney, Australia, 1995; 1.
11. Zweibel, K. *Harnessing Solar Power: The Photovoltaics Challenge*; Plenum Press: New York, 1990; 6.
12. Perlin, J. The story of solar cells. In *Organic Photovoltaics: Mechanisms, Materials and Devices*; Sun, S.-S., Sariciftci, N., Eds., Taylor and Francis: Boca Raton, FL, 2005; 9.
13. Green, M. *Silicon Solar Cells*; Centre for Photovoltaic Devices and Systems: Sydney, Australia, 1995; 2.
14. Hamakawa, Y. Background and motivation for thin-film solar cell development. *Thin-Film Solar Cells: Next Generation Photovoltaics and Its Applications*, Springer: Heidelberg, 2004; 3–4.
15. Maycock, P. World photovoltaic markets. In *Practical Handbook of Photovoltaics*; Markvart, T., Castañer, L., Eds., Elsevier: Oxford, England, 2003; 887–912.
16. Duffie, J.; Beckman, W. *Solar Engineering of Thermal Processes*; Wiley: New York, 1991; 10–11.
17. Duffie, J.; Beckman, W. *Solar Engineering of Thermal Processes*; Wiley: New York, 1991; 3–141.
18. Page, J. The role of solar radiation climatology in the design of photovoltaic systems. In *Practical Handbook of Photovoltaics*; Markvart, T., Castañer, L., Eds., Elsevier: Oxford, England, 2003; 5–66.
19. Blankenship, R. Natural organic photosynthetic solar energy transduction. In *Organic Photovoltaics: Mechanisms, Materials and Devices*; Sun, S.-S., Sariciftci, N., Eds., Taylor and Francis: Boca Raton, FL, 2005; 37–48.
20. Nelson, J. *The Physics of Solar Cells*; Imperial College Press: London, 2003; [Chapters 1–7].
21. Gray, J. The physics of the solar cell. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; [Chapter 3].
22. Hovel, H. Solar cells In *Semiconductors and Semimetals*, Vol. 11; Academic Press: Orlando, FL, 1975; [Chapters 1–4].
23. Nelson, J. *The Physics of Solar Cells*; Imperial College Press: London, 2003; 188–198.
24. Green, M. *Silicon Solar Cells*; Centre for Photovoltaic Devices and Systems: Sydney, Australia, 1995; 201–282.
25. Messenger, R.; Ventre, J. *Photovoltaic Systems Engineering*, 2nd Ed.; CRC Press: Boca Raton, FL, 2004; 373–386.
26. Tobías, I.; del Cañizo, C.; Alonso, J. Crystalline silicon solar cells and modules. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; 257.
27. Möller, H. *Semiconductors for Solar Cells*; Artech House: Norwood, MA, 1993; 4.
28. Hegedus, S.; Luque, A. Status, trends, challenges and the bright future of solar electricity from photovoltaics. In

Handbook of Photovoltaic Science and Engineering; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; 25.
29. Möller, H. *Semiconductors for Solar Cells*; Artech House: Norwood, MA, 1993; 65.
30. Nelson, J. *The Physics of Solar Cells*; Imperial College Press: London, 2003; 221–222.
31. Möller, H. *Semiconductors for Solar Cells*; Artech House: Norwood, MA, 1993; 5.
32. Saito, K.; Nishimoto, T.; Hayashi, R.; Fukae, K.; Ogawa, K. Production of a-Si:H/a-SiGe:H/a-SiGe:H stacked solar cells and their applications. *Thin-Film Solar Cells: Next Generation Photovoltaics and Its Applications*, Springer: Heidelberg, 2004; 132.
33. Deb, S. Recent advances and future opportunities for thin-film solar cells. *Thin-Film Solar Cells: Next Generation Photovoltaics and Its Applications*, Springer: Heidelberg, 2004; 19.
34. Nomoto, K.; Tomita, T. Development of amorphous-silicon single-junction solar cells and their application systems. *Thin-Film Solar Cells: Next Generation Photovoltaics and Its Applications*, Springer: Heidelberg, 2004; 113.
35. Schock, H.-W. Properties of chalcopyrite-based materials and film deposition for thin-film solar cells. *Thin-Film Solar Cells: Next Generation Photovoltaics and Its Applications*, Springer: Heidelberg, 2004; 163–178.
36. Saito, K.; Nishimoto, T.; Hayashi, R.; Fukae, K.; Ogawa, K. Production of a-Si:H/a-SiGe:H/a-SiGe:H stacked solar cells and their applications. *Thin-Film Solar Cells: Next Generation Photovoltaics and Its Applications*, Springer: Heidelberg, 2004; 121–137.
37. Lane, P.; Kafafi, Z. Solid-state organic photovoltaics: a review of molecular and polymeric devices. In *Organic Photovoltaics: Mechanisms, Materials and Devices*; Sun, S.-S., Sariciftci, N., Eds., Taylor and Francis: Boca Raton, FL, 2005; 49–97.
38. Lane, P.; Kafafi, Z. Solid-state organic photovoltaics: A review of molecular and polymeric devices. In *Organic Photovoltaics: Mechanisms, Materials and Devices*; Sun, S.-S., Sariciftci, N., Eds., Taylor and Francis: Boca Raton, FL, 2005; 50; see also page 97.
39. Deb, S. Recent advances and future opportunities for thn-film solar cells. *Thin-Film Solar Cells: Next Generation Photovoltaics and Its Applications*, Springer: Heidelberg, 2004; 37–38.
40. Nelson, J. *The Physics of Solar Cells*; Imperial College Press: London, 2003; 289–290.
41. Green, M. *Third-Generation Photovoltaics: Advanced Solar Energy Conversion*; Springer: Heidelberg, 2003; 35.
42. Green, M. *Third-Generation Photovoltaics: Advanced Solar Energy Conversion*; Springer: Heidelberg, 2003; 59–67.
43. Olson, J.; Friedman, D.; Kurtz, S. High-efficiency III–V multijunction solar cells. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; 362.
44. Green, M. *Third-Generation Photovoltaics: Advanced Solar Energy Conversion*; Springer: Heidelberg, 2003; 95–107.
45. Nelson, J. *The Physics of Solar Cells*; Imperial College Press: London, 2003; 302–309.
46. Green, M. *Third-Generation Photovoltaics: Advanced Solar Energy Conversion*; Springer: Heidelberg, 2003; 69–79.
47. Nelson, J. *The Physics of Solar Cells*; Imperial College Press: London, 2003; 309–318.
48. Green, M. *Third-Generation Photovoltaics: Advanced Solar Energy Conversion*; Springer: Heidelberg, 2003; 81–93.
49. Nelson, J. *The Physics of Solar Cells*; Imperial College Press: London, 2003; 318–323.
50. Green, M. *Third-Generation Photovoltaics: Advanced Solar Energy Conversion*; Springer: Heidelberg, 2003; 107.
51. Swanson, R. Photovoltaic concentrators. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; 449.
52. Verlinden, P. High-efficiency concentrator silicon solar cells. In *Practical Handbook of Photovoltaics*; Markvart, T., Castañer, L., Eds., Elsevier: Oxford, England, 2003; 436.
53. Jones, J. Time to concentrate. Renew. Energy **2005**, *September–October*, 78–84.
54. Slooff, L.H.; Kinderman, R.; Burgers, A.R.; et al. *The Luminescent Concentrator: A Bright Idea for Spectrum Conversion?*, Proceedings of the 20th EPVSEC, Barcelona, Spain, June 6–10 2005.
55. Messenger, R.; Ventre, J. *Photovoltaic Systems Engineering*, 2nd Ed.; CRC Press: Boca Raton, FL, 2004; 53–54.
56. Messenger, R.; Ventre, J. *Photovoltaic Systems Engineering*, 2nd Ed.; CRC Press: Boca Raton, FL, 2004; 72.
57. Lorenzo, E. Energy collected and delivered by PV modules. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; 956–962.
58. Lasnier, F.; Ang, T. *Photovoltaic Engineering Handbook*; IOP Publishing Ltd.: Bristol, England, 1990; 339–370.
59. Messenger, R.; Ventre, J. *Photovoltaic Systems Engineering*, 2nd Ed.; CRC Press: Boca Raton, FL, 2004; [Chapters 4, 7, and 8].
60. Köthe, H. Solar electric power supply with batteries.In *Battery Technology Handbook* 2nd Ed.; Kiehne, H. Ed.; Marcel Dekker: New York, 2003; [Chapter 11].
61. Markvart, T. *Solar Electricity*, 2nd Ed.; Wiley: Chichester, England, 2000; [Chapter 4].
62. Silvestre, S. Review of system design and sizing tools. In *Practical Handbook of Photovoltaics*; Markvart, T., Castañer, L., Eds., Elsevier: Oxford, England, 2003; 543–561.
63. Ross, J. System electronics. In *Practical Handbook of Photovoltaics*; Markvart, T., Castañer, L., Eds., Elsevier: Oxford, England, 2003; 565–585.
64. Schmid, J.; Schmidt, H. Power conditioning for photovoltaic power systems. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England,2003;863–902.
65. Lasnier, F.; Ang, T. *Photovoltaic Engineering Handbook*; IOP Publishing Ltd.: Bristol, England, 1990; 139–162.
66. Sauer, D. Electrochemical storage for photovoltaics. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; 799–862.
67. Vincent, C.; Scrosati, B. *Modern Batteries: An Introduction to Electrochemical Power Sources*, 2nd Ed.; Wiley: New York, 1997; 142–161.

68. Spiers, D. Batteries in PV systems. In *Practical Handbook of Photovoltaics*; Markvart, T., Castañer, L., Eds., Elsevier: Oxford, England, 2003; 589–631.
69. Lasnier, F.; Ang, T. *Photovoltaic Engineering Handbook*; IOP Publishing Ltd.: Bristol, England, 1990; 114–133.
70. Markvart, T. *Solar Electricity*, 2nd Ed.; Wiley: Chichester, England, 2000; 260–266.
71. Messenger, R.; Ventre, J. *Photovoltaic Systems Engineering*, 2nd Ed.; CRC Press: Boca Raton, FL, 2004; 145.
72. Wolfe, P. Solar-powered products. In *Practical Handbook of Photovoltaics*; Markvart, T., Castañer, L., Eds., Elsevier: Oxford, England, 2003; 771–789.
73. Preiser, K. Photovoltaic systems. In *Handbook of Photovoltaic Science and Engineering*; Luque, A., Hegedus, S., Eds., Wiley: Chichester, England, 2003; 753–785.
74. Lasnier, F.; Ang, T. *Photovoltaic Engineering Handbook*; IOP Publishing Ltd.: Bristol, England, 1990; 371–500.

Physics of Energy

Milivoje M. Kostic
Department of Mechanical Engineering, Northern Illinois University, DeKalb, Illinois, U.S.A.

Abstract
The concept and definition of energy are elaborated, as well as different forms and classification of energy are presented. Energy is a fundamental concept indivisible from matter and space, and energy exchanges or transfers are associated with all processes (or changes), thus indivisible from time. Actually, energy is "the building block" and fundamental property of matter and space, thus fundamental property of existence. Any and every material system in nature possesses energy. The structure of any matter and field is "energetic," meaning active, i.e., photon waves are traveling in space, electrons are orbiting around an atom nucleus or flowing through a conductor, atoms and molecules are in constant rotation, vibration or random thermal motion, etc. When energy is exchanged or transferred from one system to another it is manifested as work or heat. In addition, the First Law of energy conservation and the Second Law of energy degradation and entropy generation are presented along with relevant concluding remarks. In summary, energy is providing existence, and if exchanged, it has ability to perform change.

INTRODUCTION: FROM WORK TO HEAT TO GENERAL ENERGY CONCEPT

Energy is a fundamental concept indivisible from matter and space, and energy exchanges or transfers are associated with all processes (or changes), thus indivisible from time. Actually, energy is "the building block" and fundamental property of matter and space, thus a fundamental property of existence. Energy transfer is needed to produce a process to change other properties. Also, among all properties, energy is the only one which is directly related to mass and vice versa: $E = mc^2$ (known in some literature as *mass energy*, the c is the speed of light in a vacuum), thus the mass and energy are interrelated.[1,2]

Energy moves cars and trains, and boats and planes. It bakes foods and keeps them frozen for storage. Energy lights our homes and plays our music. Energy makes our bodies alive and grow, and allows our minds to think. Through centuries, people have learned how to use energy in different forms in order to do work more easily and live more comfortably. No wonder that energy is often defined as *ability to perform work*, i.e., as a potential for energy transfer in a specific direction (displacement in force direction) thus achieving a purposeful process, as opposed to dissipative (less-purposeful) energy transfer in form of heat. The above definition could be generalized as: *energy is providing existence, and if exchanged, it has the ability to perform change*.[3]

Any and every material system in nature possesses energy. The structure of any matter and field is "energetic," meaning active, i.e., photon waves are traveling in space, electrons are orbiting around an atom nucleus or flowing through a conductor, atoms and molecules are in constant rotation, vibration or random thermal motion, etc., see Table 1 and Fig. 1. Thus energy is a property of a material system (further on simply referred to as *system*), and together with other properties defines the system equilibrium state or existence in space and time.

Energy in transfer ($E_{transfer}$) is manifested as work (W) or heat (Q) when it is exchanged or transferred from one system to another, as explained next (see Fig. 2).

Work

Work is a mode of energy transfer from one acting body (or system) to another resisting body (or system), with an acting force (or its component) in the direction of motion, along a path or displacement. A body that is acting (forcing) in motion-direction in time, is doing work on another body which is resisting the motion (displacement rate) by an equal resistance force, including inertial force, in opposite direction of action. The acting body (energy source) is imparting (transferring away) its energy to the resisting body (energy sink), and the amount of energy transfer is the work done by the acting onto the resisting body, equal to the product of the force component in the motion direction multiplied with the corresponding displacement, or vice versa (force multiplied by displacement component in the force direction), see Fig. 3. If the force ($\vec{F}$) and displacement vectors ($d\vec{s} = d\vec{r}$) are not constant, then integration of differential work transfers from initial (1) to final state (2), defined by the corresponding position vectors $\vec{r}$, will be necessary, see Fig. 4.

Keywords: Energy; Entropy; Heat; Heat engine; Power; Thermal energy; Total internal energy; Work.

Table 1 Material system structure and related forces and energies

Particles	Forces	Energies
Atom nucleus	Strong and weak	Nuclear
Electron shell, or electron flow	Electromagnetic	Electrical, magnetic, electromagnetic
Atoms/molecules	Inter-atomic/molecular	Chemical
Molecules	Inertial due to random collision and, potential inter-molecular	Sensible thermal
Molecules	Potential inter-molecular	Latent thermal
Molecules	Potential inter-molecular	Mechanical elastic
System mass	Inertial and gravitational	Mechanical kinetic and gravitational potential

The work is a directional energy transfer. However, it is a scalar quantity like energy, and is distinctive from another energy transfer in form of *heat*, which is due to random motion (chaotic or random in all directions) and collisions of system molecules and their structural components.

Work transfer cannot occur without existence of the resisting body or system, nor without finite displacement in the force direction. This may not always be obvious. For example, if we are holding a heavy weight or pushing hard against a stationary wall, there will be no work done against the weight or wall (neglecting their small deformations). However, we will be doing work internally due to contraction and expansion of our muscles (thus force with displacement), and that way converting (spending) a lot of chemical energy via muscle work, then dissipating it into thermal energy and heat transfer (sweating and getting tired).

Heat

Heat is another mode of energy transfer from one system at higher temperature to another at lower temperature due to their temperature difference. Fire was civilization's first great invention, long before people could read and write, and wood was the main heat source for cooking and heating for ages. However, true physical understanding of the nature of heat was discovered rather recently, in the middle of the nineteenth century, thanks to the development of the kinetic theory of gasses. Thermal energy and heat are defined as the energy associated with the random motion of atoms and molecules. The prior concept of heat was based on the caloric theory proposed by the French chemist Antoine Lavoisier in 1789. The caloric theory defines heat as a massless, colorless, odorless, and tasteless, fluid-like substance called the *caloric* that can be transferred or "poured" from one body into another. When caloric was added to a body, its temperature increased and vice versa. The caloric theory came under attack soon after its introduction. It maintained that heat is a substance that could not be created or destroyed. Yet it was known that heat can be generated indefinitely by rubbing hands together or from mechanical energy during friction, like mixing or similar. Finally, careful experiments by James P. Joule published in 1843, quantified correlation between mechanical work and heat, and thus put the caloric theory to rest by convincing the skeptics that heat was not the caloric substance after all. Although the caloric theory was totally abandoned in the middle of the nineteenth century, it contributed greatly to the development of thermodynamics and heat transfer.

Heat may be transferred by three distinctive mechanisms: conduction, convection, and thermal radiation, see Fig. 5. Heat conduction is the transfer of thermal energy due to interaction between the more energetic particles of a substance, like atoms and molecules (thus at higher temperature), to the adjacent less energetic ones (thus at lower temperature). Heat convection is the transfer of thermal energy between a solid surface and the adjacent moving fluid, and it involves the combined effects of conduction and fluid motion. Thermal radiation is the transfer of thermal energy due to the emission of electromagnetic waves (or photons) which are products of random interactions between energetic particles of a substance (thus related to the temperature). During those interactions the electron energy level is changed, thus causing emission of photons, i.e., electromagnetic thermal radiation.

The Joule's experiments of establishing equivalency between work and heat paved the way of establishing the concept of internal thermal energy, to generalize the concept of energy, and to formulate the general law of energy conservation. The total internal energy includes all other possible but mechanical energy types or forms, including chemical and nuclear energy. This allowed extension of the well-established law of mechanical energy conservation to the general law of energy conservation, known as the First Law of Thermodynamics, which includes all possible energy forms that a system could possess, and heat and all types of work as all possible energy-transfers between the systems. The Law of Energy Conservation will be elaborated later.

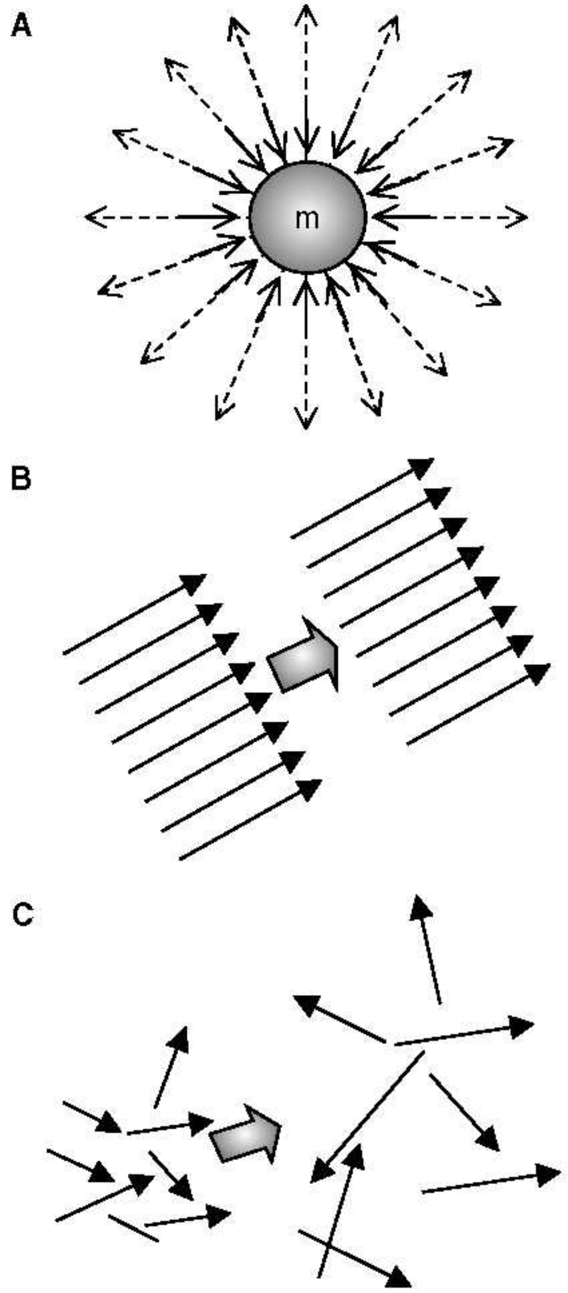


Fig. 1 Different types of energy: (A) Potential gravitational and electromagnetic radiation; (B) Organized energy as work transfer; (C) Disorganized thermal energy as heat transfer.

Energy

Energy is fundamental property of a physical system and refers to its potential to maintain a system identity or structure and to influence changes with other systems (via forced interaction) by imparting work (forced directional displacement) or heat (forced chaotic displacement/motion of a system molecular or related structures). Energy exists in many forms: electromagnetic (including light), electrical, magnetic, nuclear, chemical, thermal, and mechanical (including kinetic, elastic, gravitational, and sound); where, for example, electro-mechanical energy may be kinetic or potential, while thermal energy represents overall potential and chaotic motion energy of molecules and/or related micro structure.[3,4]

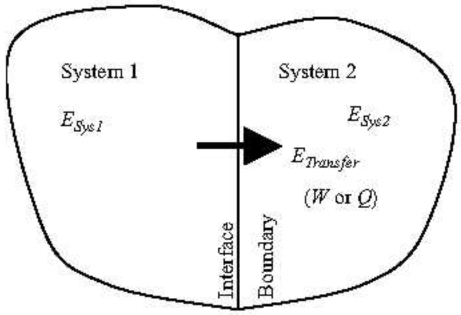


Fig. 2 Energy as material system property and energy transfer from one system to another.

Energy, Work, and Heat Units

Energy, Work, and Heat Units. Energy is manifested via work and heat transfer, with corresponding Force × Length dimension for work (N m, kg_fm, and lb_fft, in SI, metric and English system of units, respectively); and the caloric units, in kilocalorie (kcal) or British-thermal-unit (Btu), the last two defined as heat needed to increase a unit mass of water (at specified pressure and temperature) for one degree of temperature in their respective units. Therefore, the water specific heat is 1 kcal/(kg °C) = 1 Btu/(lb °F) by definition, in metric and English system of units, respectively. It was demonstrated by Joule that 4187 N m of work, when dissipated in heat, is equivalent to 1 kcal. In his honor, 1 N m of work is named after him as 1 Joule, or 1 J, the SI energy unit, also equal to electrical work of 1 W s = 1 V A s. The SI unit for power, or work rate, is watt, i.e., 1 J/s = 1 W, and also corresponding units in other system of units, like Btu/h, etc. The Horse Power is defined as 1 HP = 550 lb_fft/s = 745.7 W. Other common units for energy, work and heat are given in Table 2.

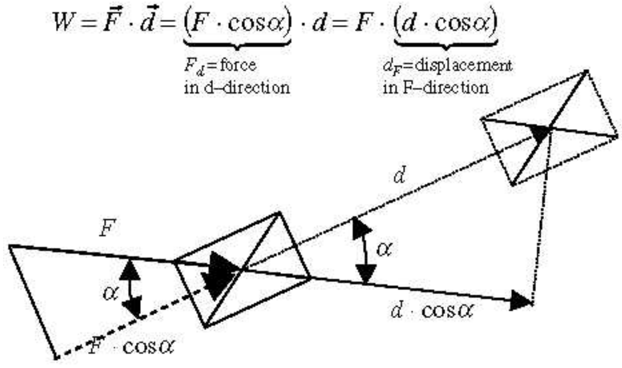


Fig. 3 Work, force, and displacement.

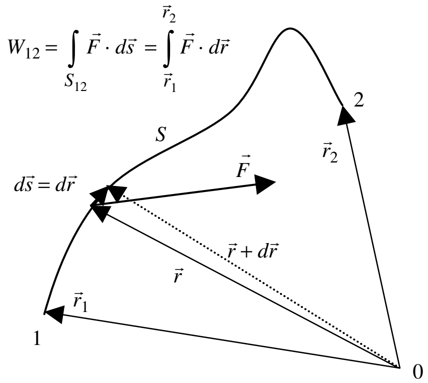


Fig. 4 Work along an arbitrary path.

ENERGY FORMS AND CLASSIFICATIONS: ENERGY-TRANSFER VS ENERGY-PROPERTY

Any and all changes or processes (happening in space and time) are caused by energy exchanges or transfers from one substance (*system* or *subsystem*) to another, see Fig. 2. A part of a system may be considered as a subsystem if energy transfer within a system is taking place, and inversely, a group of interacting systems may be considered as a larger isolated system, if they do not interact with the rest of the surroundings. Energy transfer may be in organized form (different types of work transfer due to different force actions) or in chaotic disorganized form (heat transfer due to temperature difference). Energy transfer into a system builds up energy–potential or generalized-force (called simply potential for short, like pressure, temperature, voltage, etc.) over energy—displacement (like volume, entropy, charge, etc.). Conversely, if energy is transferred from a system, its energy potential is decreased. That is why net-energy is transferred from higher to lower energy potential only, due to virtually infinite probability of equi-partition of energy over system micro-structure, causing system equilibrium, otherwise virtually impossible singularity of energy potential at infinite magnitude would result.[4]

There are many forms and classifications of energy, see Table 3, all of which could be classified as *microscopic* (or internal within a system microstructure) and/or *macroscopic* (or external as related to the system mass as a whole with reference to other systems). Furthermore, energy may be "quasi-potential" (associated with a system equilibrium state and structure, i.e., system property) or "quasi-kinetic" (energy in-transfer from one system or one structure to another, in form of *work* or *heat*).

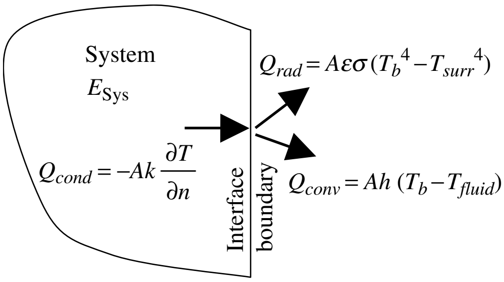


Fig. 5 Heat as energy-transfer by conduction, convection, and radiation is due to a difference in temperature.

Every material system state is an equilibrium potential "held" by forces (space force-fields), i.e., the forces "define" the potential and state—action and reaction; otherwise a system will undergo dynamic change (in time), or quasi-kinetic energy exchange towards another stable equilibrium. Atoms (mass) are "held" by atomic and electromagnetic forces in small scale and by gravity in large scale, see Fig. 1A, otherwise mass would disintegrate ("evaporate" or radiate into energy) like partly in nuclear reactions—*nuclear* energy or electromagnetic radiation. Molecules are "held" by electro-chemical bonding (valence) forces (chemical reactions—*chemical* energy). Liquids are "held" by latent intermolecular forces (gas condensation, when kinetic energy is reduced by cooling—*latent thermal* energy). Solids are "held" by "firm" intermolecular forces (freezing/solidification when energy is further reduced by cooling—*latent thermal* energy again). *Sensible thermal* energy represents intermolecular potential energy and energy of random molecular motion, and is related to temperature of a system. "Holding" potential forces may be "broken" by energy transfer (e.g., radiation, heating, high-energy particles interactions, etc.). States and potentials are often "hooked" (i.e., stable) and thus need to be "unhooked" (or to "be broken") to overcome existing "threshold" or equilibrium, like in igniting combustion, starting nuclear reaction, etc.

As stated above, energy transfer can be directional (purposeful or organized) and chaotic (dissipative or disorganized). For example, mass-in-motion, *mechanical kinetic* energy, and electricity-in-motion, *electrical kinetic* energy, are organized kinetic energies (Fig. 1B), while *thermal* energy is disorganized chaotic energy of motion of molecules and atoms (Fig. 1C). System energy may be defined with reference to position in a vector-force field, like *elastic potential* (stress) energy, *gravitational potential* energy, or *electromagnetic* field energy. There are many different energy forms and types (see Table 3). We are usually not interested in (absolute) energy level, but in the change of energy (during a process) from an initial state (i) to a final state (f), and thus zero reference values for different energy forms are irrelevant, and often taken arbitrarily for convenience. The followings are basic correlations for energy changes of several typical energy forms, often encountered in practice: motion kinetic energy ($E_K = KE$) as a function of system velocity (V); potential elastic-deformational, e.g., pressure elastic or spring elastic energy ($E_{Pdeff} = E_{Pp} = E_{Ps}$) as a function of spring deformation displacement (x); gravitational potential energy ($E_{Pg} = PE_g$) as a function of gravitational

Table 2 Typical energy units with conversion factors

Energy units	J	kWh	Btu
1 Joule (J)	1	2.78×10^{-7}	9.49×10^{-4}
1 Kilowatt-hour (kWh)	3.6×10^{6}	1	3.412×10^{3}
1 Kilocalorie (kcal = Cal = 1000 cal)	4187	1.19×10^{-3}	3.968
1 British thermal unit (Btu)	1055	2.93×10^{-4}	1
1 Pound-force foot ($lb_f\cdot ft$)	1.36	3.78×10^{-7}	1.29×10^{-3}
1 Electron volt (eV)	1.6×10^{-19}	4.45×10^{-26}	1.52×10^{-22}
1 Horse Power × second (HP sec)	745.7	2.071×10^{-4}	0.707

elevation (z); and sensible thermal energy ($E_U = U$) as a function of system temperature (T):

$$\Delta E_k = \frac{1}{2} m\left(V_f^2 - V_i^2\right); \qquad \Delta E_{Ps} = \frac{1}{2} k\left(x_f^2 - x_i^2\right)$$
$$\Delta E_{Pg} = mg(z_f - z_i); \qquad \Delta E_U = mc_v(T_f - T_i) \tag{1}$$

If the reference energy values are taken to be zero when the above initial (i) variables are zero, then the above equations will represent the energy values for the final values (f) of the corresponding variables. If the corresponding parameters, spring constant k, gravity g, or constant-volume specific heat c_v, are not constant, then integration of differential energy changes from initial to final state will be necessary.

Energy transfer via work W (net-out), and heat transfer Q (net-in), may be expressed for reversible processes as product of related energy–potentials (pressure P, or temperature T) and corresponding energy–displacements (change of volume V and entropy S, respectively), i.e.:

$$W_{12} = \vec{F}\cdot\vec{d}$$
$$= (P\cdot \underbrace{A\vec{n})\cdot\vec{d}}_{\Delta V} = P\cdot\Delta V_{12}\big|_{P\neq\text{const}} = \int_{V_1}^{V_2} P\mathrm{d}V \tag{2}$$

$$Q_{12} = T\Delta S_{12}\big|_{T\neq\text{const}} = \int_{S_1}^{S_2} T\mathrm{d}S \tag{3}$$

Note, in Eq. 2, that force cannot act at a point but is distributed as pressure (P) over some area A (with orthogonal unit vector $\vec{n}$), which when displaced will cause the volume change ΔV. Also note that it is custom in some references to denote heat transfer "in" and work transfer "out" as positive (as they appear in a heat engine). In general, "in" (means "net-in") is negative "out" (means "net-out") and vice versa.

In general, energy transfer between systems is taking place at the system boundary interface and is equal to the product of energy–potential or generalized-force and the corresponding generalized-displacement[5]:

$$\begin{aligned}\delta E_{\text{Transfer}} &= \delta Q_{\text{netIN}} - \left[\sum \delta W_{\text{netOUT}}\right] \\ &= \delta Q_{\text{netIN}} + \left[\sum \delta W_{\text{netIN}}\right] \\ &= T\mathrm{d}S + \Bigg[\underbrace{-P\mathrm{d}V}_{\text{COMPR.}} + \underbrace{\sigma\mathrm{d}A}_{\text{STRECHING}} \\ &\quad + \underbrace{\tau\mathrm{d}(As)}_{\text{SHEARING}} + \underbrace{V\mathrm{d}q}_{\text{CHARGING}} + \underbrace{\vec{E}\mathrm{d}(V\vec{P})}_{\text{POLARIZATION}} \\ &\quad + \underbrace{\mu_o\vec{H}\mathrm{d}(V\vec{M})}_{\text{MAGNETIZATION}} + \underbrace{\quad}_{\text{ETC}}\Bigg]\end{aligned} \tag{4}$$

Where the quantities after the last equal sign are: temperature and entropy; pressure and volume; surface tension and area; tangential-stress and area with tangential-displacement, voltage and electrical charge; electric field strength and volume-electric dipole moment per unit volume product; and permeability of free space, magnetic field strength and volume-magnetic dipole moment per unit volume product; respectively.

The total system energy stored within the system, as energy property, is:

$$\begin{aligned}E_{\text{Sys}} &= \underbrace{E_K + E_{Pg} + E_{\text{Pdeff.}}}_{E_{\text{Mechanical}}} \\ &\quad \underbrace{+ \underbrace{E_{\text{Uth}}}_{\text{Thermal}} + E_{\text{Ch}} + E_{\text{Nucl}} + E_{\text{El}} + E_{\text{Magn}} + \underbrace{\ldots}_{\text{Etc.}}}_{\text{Internal(total)}}\end{aligned} \tag{5}$$

Where the quantities after the equal sign are: kinetic, potential-gravitational, potential-elastic-deformational, thermal, chemical, nuclear, electric, magnetic energies, etc.

THE FIRST LAW OF ENERGY CONSERVATION: WORK–HEAT–ENERGY PRINCIPLE

Newton formulated the general theory of motion of objects due to applied forces (1687). This provided for

Table 3 Energy forms and classifications

ENERGY									
Scale					*Type*			*Transfer (release)*	
						Kinetic (motion)		**Kinetic (motion)**	
Macro/external (mass-based)	**Micro/internal (structure-based)**	*Energy form*	*(Energy storage)*	*Energy process*	**Potential (state or field)**	**Directional[a]**	**Chaotic dissipative**	**Work[b] directional[a]**	**Heat dissipative**
		Mechanical							
X		Kinetic	$mV^2/2$	Acceleration		X		X	
X		Gravitational[c]	mgz	Elevation	X			X	
	X	Elastic	$kx^2/2$ or $PV=mP/\rho$	Deformation	X			X	
		Thermal	U_{th}						
	X	Sensible	$U_{th}=mc_{avg}T$	Heating			X		X
	X	Latent	$U_{th}=H_{latent}$	Melting, Evaporation	X				X
	X	Chemical	U_{ch}	Chemical Reaction	X				X
	X	Nuclear	U_{nucl}	Nuclear Reaction	X				X
		Electrical	E_{el}						
	X	Electro-kinetic	$V(It)$ or $LI^2/2$	Electro-current flow		X		X	
	X	Electrostatic	$(It)^2/(2C)$	Electro-charging	X			X	
		Magnetic	E_{magn}	Magnetization	X			X	
	X[d]	Electromagnetic	E_{el_mag}	Radiation		X[d]		X	

[a]Electro-mechanical kinetic energy type (directional/organized, the highest energy quality) is preferable since it may be converted to any other energy form/type with high efficiency.
[b]All processes (involving energy transfer) are to some degree irreversible (i.e., dissipative or chaotic/disorganized).
[c]Due to mass position in a gravitational field.
[d]Electromagnetic form of energy is the smallest known scale of energy.

concepts of mechanical work, kinetic and potential energies, and development of solid-body mechanics. Furthermore, in absence of non-mechanical energy interactions, excluding friction and other dissipation effects, it is straightforward to derive (and thus prove) energy conservation, i.e.:

$$\underbrace{\int_{s_1}^{s_2} F_s \mathrm{d}s}_{W_{Fs}} = \int_{s_1}^{s_2}\left(m\underbrace{\frac{\mathrm{d}V_s}{\mathrm{d}t}}_{a_s}\right)\mathrm{d}s = \int_{s_1}^{s_2}\left(m\frac{\mathrm{d}V_s}{\mathrm{d}s}\underbrace{\left\{\frac{\mathrm{d}s}{\mathrm{d}t}\right\}}_{V_s}\right)\mathrm{d}s$$
$$= \int_{V_1}^{V_2} mV_s \mathrm{d}V_s = \underbrace{\frac{1}{2}m\left(V_2^2 - V_1^2\right)}_{KE_2 - KE_1} \tag{6}$$
$$\left(E_{Transfer} = W_{Fs}\right) = \left(KE_2 - KE_1 = E_2 - E_1\right)$$

The above correlation is known as the *work–energy principle*. The work–energy principle could be easily expended to include work of gravity force and gravitational potential energy as well as elastic spring force and potential elastic spring energy.[3]

During a free gravity fall (or free bounce) without air friction, for example, the potential energy is being converted to kinetic energy of the falling body (or vice versa for free bounce), and at any time the total mechanical energy (sum of kinetic and potential mechanical energies) is conserved, i.e., stays the same, see Fig. 6. The mechanical energy is also conserved if a mass freely vibrates on an ideally elastic spring, or if a pendulum oscillates around its pivot, both in absence of dissipative effects, like friction or non-elastic deformation. In general, for work of conservative forces only, the mechanical energy, E_{mech} , for N isolated systems, is conserved since

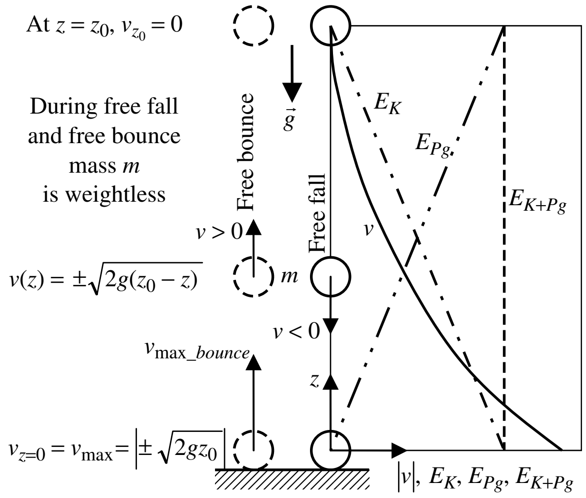


Fig. 6 Energy and work due to conservative gravity force.

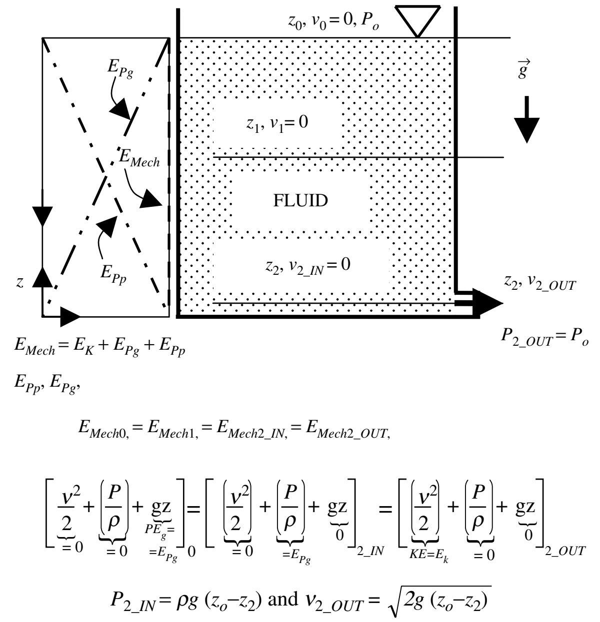


Fig. 7 Conservation of fluid mechanical energy: Bernoulli equation, hydrostatic equation, and Torricelli's orifice velocity.

there is no dissipative conversion into thermal energy and thus no heat transfer, i.e.:

$$
\begin{aligned}
E_{\text{mech}} &= E_{\text{k}} + E_{\text{Pg}} + E_{\text{Ps}} \\
&= \sum_{i=1}^{N} \left(\frac{1}{2} mV^2 + mgz + \frac{1}{2} kx^2 \right)_i \\
&= \text{const}
\end{aligned}
\tag{7}
$$

The mechanical work–energy concept could also be expended to fluid motion by inclusion of elastic-pressure force and potential elastic-pressure energy (referred to in some references as flow work; however, note that elastic-pressure energy is a system property while the flow work is related energy transfer), see the Bernoulli equation below. For flowing or stationary fluid without frictional effects, the mechanical energy is conserved, including fluid elastic—pressure energy, PV $= mP/\rho$ (where V is volume, whereas v is used for velocity here), as expressed by the Bernoulli or hydrostatic equations below, see also Fig. 7.[3]

$$
\begin{aligned}
\frac{E_{\text{mech}}}{m} &= \frac{1}{m}\left(E_{\text{K}} + E_{\text{Ps}} + E_{\text{Pg}}\right) \\
&= \underbrace{\frac{v^2}{2} + \frac{P}{\rho} + gz}_{\text{Bernoulli equation}} \Bigg|_{v=0} \\
&= \underbrace{\frac{P}{\rho} + gz}_{\text{Hydrostatic equation } (v=0)} = \text{const.}
\end{aligned}
\tag{8}
$$

Work against inertial and/or conservative forces (also known as internal, or volumetric, or space potential field), is path-independent and during such a process the mechanical energy is conserved. However, work of non-conservative, dissipative forces is process path-dependent and part of the mechanical energy is converted (dissipated) to thermal energy, see Fig. 8A.

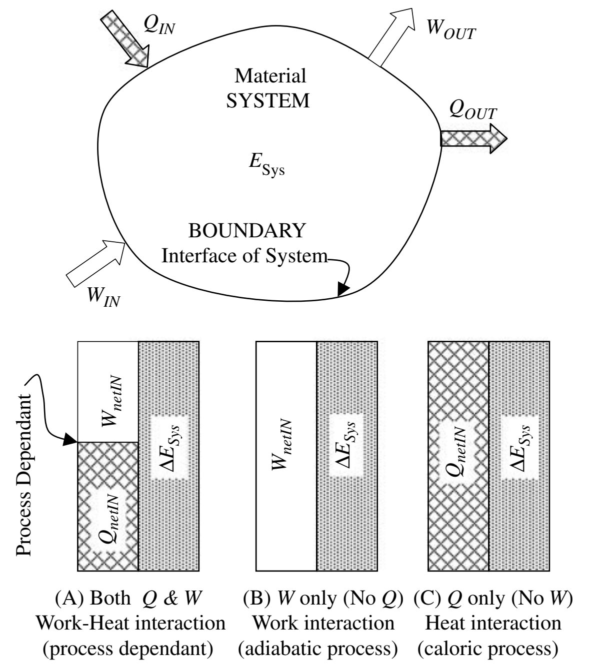


Fig. 8 System energy and energy boundary interactions (transfers) for (A) arbitrary, (B) adiabatic, and (C) caloric processes.

When work of non-conservative forces W_{nc} , is exchanged between N isolated systems, from an initial (i) to final state (f), then the total mechanical energy of all systems is reduced by that work amount, i.e.:

$$W_{nc,i \to f} = \left(\sum_{j=1}^{N} E_{mech,j} \right)_i - \left(\sum_{j=1}^{N} E_{mech,j} \right)_f \tag{9}$$

Regardless of the traveled path (or displacement), the work against conservative forces (like gravity or elastic spring in above cases) in absence of any dissipative forces, will depend on the final and initial position (or state) only. However, the work of non-conservative, dissipative forces (W_{nc}) depends on the traveled path since the energy is dissipated during the force displacement, and mechanical energy will not be conserved, but in part converted (via dissipation and heat transfer) into thermal energy, see Eq. 9. This should not be misunderstood with the total energy conservation, which is always the case, and it must include both work and heat transfer, see below.

As already stated, there are many different types of work transfer into (or out of) a system which will change the corresponding energy-form stored in (or released, discharged out of) the system. In addition to work, energy may be transferred as heat caused by temperature difference and associated with change of the thermal energy of a system. Furthermore, different forms of stored energy are often coupled so that one type of energy transfer may change more than one form of stored energy, particularly due to inevitable dissipative conversion of work to heat, and in turn to internal thermal energy. Conversely, heat and thermal energy may be converted into other energy forms. In the absence of nuclear reaction (no conversion of mass into energy, $E=mc^2$), mass and energy are conserved separately for an isolated system, a group of isolated systems, or for the Universe. Since the material system structure is of particulate form, then systems' interactions (collisions at different scale-sizes) will exchange energy during the forced displacement—and similarly to the mechanical energy conservation—the totality of all forms of energy will be conserved,[8] see Fig. 8, which could be expressed as:

$$\underbrace{\sum_{\text{All } i\text{'s}} E_{i,\text{Trans.}}}_{\text{BOUNDARY}} = \underbrace{\Delta E_{\text{Change}}}_{\text{SYSTEM}} \quad \text{or}$$

$$\underbrace{\sum_{\text{All } j\text{'s}} W_{j,\text{netIN}} + \sum_{\text{All } k\text{'s}} Q_{k,\text{netIN}}}_{\text{BOUNDARY}} = \underbrace{\Delta E_{\text{netIncrease}}}_{\text{SYSTEM}} = \Delta E_{\text{Sys}} \tag{10}$$

Energy interactions or transfers across a system boundary, in the form of work, $W_{netIN} = \sum W_{IN} - \sum W_{OUT} = -\left(\sum W_{OUT} - \sum W_{IN}\right) = -W_{netOUT}$, and heat, $Q_{netIN} = \sum Q_{IN} - \sum Q_{OUT}$, will change the system total energy, $\Delta E_{Sys} = E_{Sys,2} - E_{Sys,1}$. The boundary energy transfers are process (or process path) dependent for the same ΔE_{Sys} change, except for special cases for adiabatic processes with work interaction only (no heat transfer), or for caloric processes with heat interaction only (no work transfer), see Fig. 8A to C. For the latter caloric processes without work interactions (no volumetric expansion or other mechanical energy changes), the internal thermal energy is conserved by being transferred from one system to another via heat transfer only, known as *caloric*. This demonstrates the value of the caloric theory of heat that was established by Lavoisier and Laplace (1789), the great minds of the 18th century. Ironically, the caloric theory was creatively used by Sadi Carnot to develop the concept of reversible cycles for conversion of caloric heat to mechanical work as it "flows" from high to low temperature reservoirs (1824) that later helped in dismantling the caloric theory. The caloric theory was discredited by establishing the *heat equivalent of work*, e.g., *mechanical equivalent of heat* by Mayer (1842) and experimentally confirmed by Joule (1843), which paved the way for establishing the First Law of Energy Conservation and new science of Thermodynamics (Clausius, Rankine, and Kelvin, 1850 and later). Prejudging the caloric theory now as a "failure" is unfair and unjustified since it made great contributions in calorimetry and heat transfer, and it is valid for caloric processes (without work interactions). The coupling of work–heat interactions and conversion between thermal and mechanical energy are outside of the caloric theory domain and are further developed within *the First* and *the Second Laws* of Thermodynamics.

The First Law of energy conservation for the control-volume (CV, with boundary surface BS) flow process, see Fig. 9, is:

$$\begin{aligned}
&\underbrace{\frac{\mathrm{d}}{\mathrm{d}t} E_{CV}}_{\text{RATE OF ENERGY CHANGE IN CV}} \\
&= \underbrace{\sum_{BS} \dot{Q}_{netIN,i}}_{\text{BS TRANSFER RATE OF HEAT}} \\
&- \underbrace{\sum_{BS} \dot{W}_{netOUT,i}}_{\text{BS TRANSFER RATE OF WORK}} \\
&+ \underbrace{\sum_{IN} \dot{m}_j (e + Pv)_j}_{\text{ENERGY TRANSPORT RATE WITH MASS IN}} \\
&- \underbrace{\sum_{OUT} \dot{m}_k (e + Pv)_k}_{\text{ENERGY TRANSPORT RATE WITH MASS OUT}}
\end{aligned} \tag{11}$$

The First Law of energy conservation equation for a differential volume per unit of volume around a point

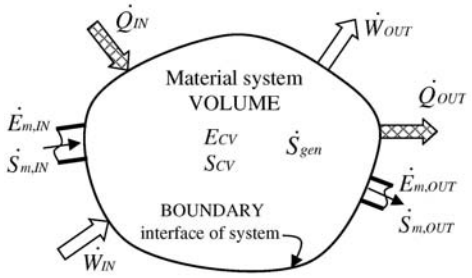


Fig. 9 Control-volume (CV) energy and entropy, and energy and entropy flows through the boundary interface of the control volume.

(x,y,z) in a flowing fluid is[6]:

$$\underbrace{\rho\frac{De}{Dt}}_{\text{energy change in time}} = \underbrace{-\vec{V}\cdot(\nabla P)}_{\text{work rate of normal pressure stresses}} + \underbrace{\nabla\cdot\left(\vec{V}\cdot\tau_{ij}\right)}_{\text{work rate of shearing stresses}} + \underbrace{\nabla\cdot(k\nabla T)}_{\text{heat rate via thermal conduction}} \quad (12)$$

Where $e = \hat{u} + \frac{\vec{V}^2}{2} + gz$ NOTE: distinguish specific thermal energy $\hat{u}$, from velocity component u below.

Eq. 12 after substitution, $\nabla\cdot\left(\vec{V}\cdot\tau_{ij}\right) = \vec{V}\cdot\left(\nabla\cdot\tau_{ij}\right) + \Phi$, and using the momentum equation, reduces to:

$$\rho\frac{D\hat{u}}{Dt} = -p\left(\nabla\cdot\vec{V}\right) + \Phi_\kappa + \Phi + \nabla\cdot(k\nabla T) \quad (13)$$

Where, $\Phi_\kappa = \kappa\left(\nabla\cdot\vec{V}\right)^2$ is the bulk-viscosity dissipation, and Φ is the shear-viscosity dissipation function, which is the rate of mechanical work conversion to internal thermal energy for a differential volume per unit of volume, with [W/m^3] unit, is given for Newtonian fluid as:

$$\Phi = \frac{d\dot{W}_\Phi}{dV} = \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial w}{\partial z}\right)^2\right] + \mu\left[\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}\right)^2 + \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right)^2\right] - \frac{2}{3}\mu\left(\nabla\cdot\vec{V}\right)^2 \quad (14)$$

The power or work rate of viscous dissipation (irreversible conversion of mechanical to thermal energy) in a control volume V is:

$$\dot{W}_\Phi = \int_V \Phi \, dV \quad (15)$$

NOTE : distinguish volume V, form velocity $\vec{V}$

THE SECOND LAW OF ENERGY DEGRADATION: ENTROPY AND EXERGY

Every organized kinetic energy will, in part or in whole (and ultimately in whole), disorganize/dissipate within the microstructure of a system (in time over its mass and space) into disorganized thermal energy. Entropy, as an energy–displacement system property, represents the measure of energy disorganization, or random energy redistribution to smaller-scale structure and space, per absolute temperature level. Contrary to energy and mass, which are conserved in the universe, entropy is continuously generated (increased) due to continuous redistribution and disorganization of energy in transfer and thus degradation of the quality of energy ("spreading" of energy towards and over lower potentials in time until equi-partitioned over system structure and space). Often, we want to extract energy from one system in order to purposefully change another system, thus to transfer energy in organized form (as work, thus the ultimate energy quality). No wonder that energy is defined as *ability to perform work*, and a special quantity *exergy* is defined as the maximum possible work that may be obtained from a system by bringing it to the equilibrium in a process with reference surroundings (called *dead state*). The maximum possible work will be obtained if we prevent energy disorganization, thus with limiting *reversible* processes where the existing non-equilibrium is conserved within interacting systems. Since the energy is conserved during any process, only in ideal reversible processes entropy (measure of energy disorganization or degradation) and exergy (maximum possible work with regard to the reference surroundings) will be conserved, while in real irreversible processes, the entropy will be generated and exergy will be partly (or even fully) destroyed. Therefore, heat transfer and thermal energy are universal manifestations of all natural and artificial (man-made) processes, where all organized potential and/ or quasi-kinetic energies are disorganized or dissipated in the form of thermal energy, in irreversible and spontaneous processes.

Reversibility and Irreversibility: Energy Transfer and Disorganization, and Entropy Generation

Energy transfer (when energy moves from one system or subsystem to another) through a system boundary and in time, is of kinetic nature, and may be directionally organized as work or directionally chaotic and disorganized as heat. However, the net-energy transfer is in one direction only, from higher to lower energy–potential, and the process cannot be reversed. Thus *all processes are irreversible* in the direction of decreasing energy–potential (like pressure and temperature) and increasing energy–displacement (like volume and entropy) as a consequence of energy and mass conservation in the universe. This implies that the universe (as isolated and unrestricted system) is expending in space with entropy generation (or increase) as a measure of continuous energy degradation, i.e., energy redistribution and disorganization. It is possible in the limit to have an energy transfer process with infinitesimal potential difference (still from higher to infinitesimally lower potential, P). Then, if infinitesimal change of potential difference direction is reversed ($P + dP \rightarrow P - dP$, with infinitesimally small external energy, since $dP \rightarrow 0$), the process will be reversed too, which is characterized with infinitesimal entropy generation, thus in the limit, without energy degradation (no further energy disorganization) and no entropy generation—thus achieving a limiting, ideal *reversible process*. Such processes at infinitesimal potential differences, called quasiequilibrium processes, maintain the system equilibrium at any instant but with incremental changes in time. Only quasiequilibrium processes are reversible and vice versa. In effect, the quasiequilibrium reversible processes are infinitely slow processes at infinitesimally small potential differences, but they could be reversed to any previous state, and forwarded to any future equilibrium state, without any "permanent" change to the rest of the surroundings. Therefore, the changes are "fully" reversible, and along with their rate of change and time, completely irrelevant, as if nothing is effectively changing—the time is irrelevant as if it does not exist since it could be reversed (no permanent change and relativity of time). Since the real time cannot be reversed, it is a measure of permanent changes, i.e., irreversibility, which is in turn measured by entropy generation. In this regard the time and irreversible entropy generation are related.[2]

Entropy is also a system property, which together with energy defines its equilibrium state, and actually represents the system energy–displacement or random energy disorganization (dissipation) per absolute temperature level over its mass and space it occupies. Therefore, the entropy as property of a system, for a given system state, is the same regardless whether it is reached by reversible heat transfer, Eq. 16, or irreversible heat or irreversible work transfer (adiabatic or caloric processes on Fig. 8B and C). For example, an ideal gas system entropy increase will be the same during a reversible isothermal heat transfer and reversible expansion to a lower pressure (heat-in equal to expansion work-out), as during an irreversible adiabatic unrestricted expansion (no heat transfer and no expansion work) to the same pressure and volume, as illustrated in Fig. 10A and B, respectively.

If heat or work at higher potential (temperature or pressure) than necessary, is transferred to a system, the energy at excess potential will dissipate spontaneously to a lower potential (if left alone) before a new equilibrium state is reached, with entropy generation, i.e., increase of entropy (energy degradation per absolute temperature level). A system will "accept" energy if it is transferred at minimum necessary (infinitesimally higher) or higher potential with regard to the system energy–potential. Furthermore, the higher potential energy will dissipate and entropy increase will be the same as with minimum necessary potential, like in reversible heating process, i.e.:

$$dS = \frac{\delta Q}{T} \quad \text{or} \quad S = \int \frac{\delta Q}{T} + S_{ref} \tag{16}$$

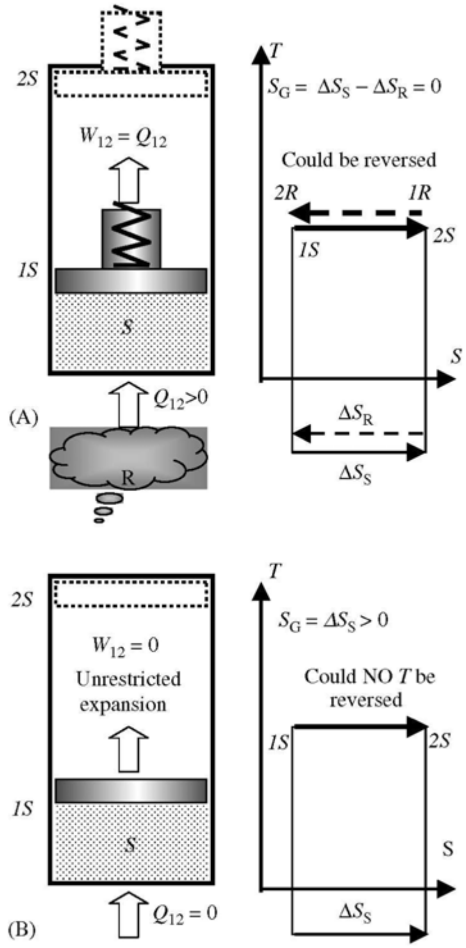


Fig. 10 (A) Isothermal reversible heat transfer and restricted reversible expansion; (B) Adiabatic unrestricted irreversible expansion of the same initial system to the same final state (Thus the same system entropy change).

However, the source entropy will decrease to a smaller extent over higher potential, thus resulting in overall entropy generation for the two (or all) interacting systems, which may be considered as a combined isolated system (no energy exchange with the rest of the surroundings). The same is true for energy exchange between different system parts (could be considered as subsystems) at different energy potentials (non uniform, not at equilibrium at a given time). Energy at higher potential (say close to boundary within a system) will dissipate ("mix") to parts at lower energy potential with larger entropy increase than decrease at higher potential, resulting in internal irreversibility and entropy generation, i.e., energy "expansion" over more mass and/or space with lower potential. Entropy is not displacement for heat only as often stated, but also displacement for any energy dissipation (energy disorganization) and the measure of irreversibility. Examples are unrestricted or throttling expansion with no heat exchange but entropy generation. Therefore, entropy generation is fundamental measure of irreversibility or "permanent change."

Even though directionally organized energy transfer as work, does not possess or generate any entropy (no energy disorganization, Fig. 1B), it is possible to obtain work from the equal amount of disorganized thermal energy or heat if such process is reversible. There are two typical reversible processes where disorganized heat or thermal energy could be entirely transferred into organized work, and vice versa. Namely, they are[2]:

1. reversible expansion at constant thermal energy, e.g., isothermal ideal-gas expansion ($\delta W = \delta Q$), Fig. 10A, and
2. reversible adiabatic expansion ($\delta W = -\mathrm{d}U$).

During a reversible isothermal heat transfer and expansion of an ideal-gas system (S), for example, Fig. 10A, the heat transferred from a thermal reservoir (R) will reduce its entropy for (ΔS_R magnitude), wile ideal gas expansion in space (larger volume and lower pressure) will further disorganize its internal thermal energy and increase the gas entropy for (ΔS_S), while in the process an organized expansion work, equivalent to the heat transferred, will be obtained ($W_{12} = Q_{12}$). The process could be reversed, and thus it is reversible process with zero total entropy generation ($S_G = \Delta S_S - \Delta S_R = 0$). On the other hand, if the same initial system (ideal gas) is expanded without any restriction (Fig. 10B, thus zero expansion work) to the same final state, but without heat transfer, the system internal energy will remain the same but more disorganized over the larger volume, resulting in the same entropy increase as during the reversible isothermal heating and restricted expansion. However, this process can not be reversed, since no work was obtained to compress back the system, and indeed the system entropy increase represents the total entropy generation ($S_G = \Delta S_S > 0$). Similarly, during reversible adiabatic expansion, the system internal thermal energy will be reduced and transferred in organized expansion work with no change of system entropy (isentropic process), since the reduction of disorganized internal energy and potential reduction of entropy will be compensated with equal increase of disorder and entropy in expanded volume. The process could be reversed back-and-forth (like elastic compression–expansion oscillations of a system) without energy degradation and entropy change, thus an isentropic processes. In reversible processes energy is exchanged at minimum-needed, not higher than needed potential, and isolated systems do not undergo any energy–potential related degradation/disorganization, and with total conservation of entropy. The total non-equilibrium is conserved by reversible energy transfer within interacting systems, i.e., during reversible processes.

There are two classical statements of the Second Law (both negative, about impossibility), see Fig. 11.[7,8] One is the *Kelvin–Plank statement* which expresses the never violated fact that it is impossible to obtain work in a continuous cyclic process (*perpetual mobile*) from a single thermal reservoir (100% efficiency impossible, since it is not possible to spontaneously create non-equilibrium

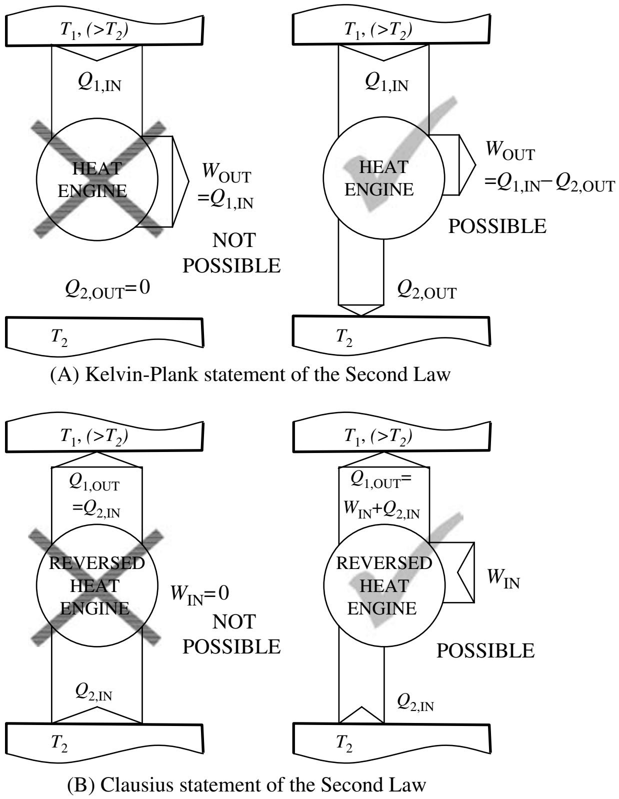


Fig. 11 The Second Law: (A) Kelvin–Plank, and (B) Clausius statements.

within the single reservoir equilibrium), and the second *Clausius statement* refers to the direction of heat transfer, expressing the never violated fact that it is impossible for heat transfer to take place by itself (without any work input) from a lower to higher temperature thermal reservoir (it is impossible to spontaneously create non-equilibrium). Actually the two statements imply each other and thus are the same, as well as they imply that all reversible cycles between the two temperature reservoirs (or all reversible processes between the two states) are the most and equally efficient with regards to extracting the maximum work out of a system, or requiring the minimum possible work into the system (thus conserving the existing non-equilibrium). A heat-engine cycle, energy conversion efficiency is defined as:

$$\begin{aligned}\eta_{\text{cycle}} &= \frac{W_{\text{cycle,netOUT}}}{Q_{\text{IN}}} = \frac{Q_{\text{IN}} - Q_{\text{OUT}}}{Q_{\text{IN}}} \\ &= 1 - \left.\frac{Q_{\text{OUT}}}{Q_{\text{IN}}}\right|_{\text{rev}} \\ &= 1 - \frac{T_{\text{OUT}}}{T_{\text{IN}}} = \eta_{\text{rev}} = \eta_{\text{max}}\end{aligned} \tag{17}$$

The reversible cycle efficiency between the two thermal reservoirs does not depend on the cycling system, but only on the ratio of the absolute temperatures of the two reservoirs ($T{=}T_{\text{IN}}$ and $T_{\text{ref}}{=}T_{\text{OUT}}$), known as *Carnot* cycle efficiency (last part in the Eq. 17, see also below). The latter defines the *thermodynamic temperature* with the following correlation:

$$\frac{T}{T_{\text{ref}}} = \left.\frac{Q}{Q_{\text{ref}}}\right|_{\text{rev cycle}} \tag{18}$$

The Second Law efficiency is defined by comparing the real irreversible processes or cycles with the corresponding ideal reversible processes or cycles:

$$\underbrace{\eta_{\text{II,OUT}} = \frac{W_{\text{OUT}}}{W_{\text{OUT,rev}}}}_{\text{Energy Production Process}} \quad \text{or} \quad \underbrace{\eta_{\text{II,IN}} = \frac{W_{\text{IN,rev}}}{W_{\text{IN}}}}_{\text{Energy Consumption Process}} \tag{19}$$

The irreversibility (I) is due to the entropy generation (S_{gen}) and represents the lost work potential or lost exergy ($E_{\text{X,loss}}$) with regard to reference system (surroundings at T_{o} absolute temperature), as expressed by the following correlation:

$$I = E_{\text{X,loss}} = T_{\text{o}} S_{\text{gen}} \tag{20}$$

The entropy balance equation for the control-volume flow process, complementing the related energy balance Eq. 11, see Fig. 9, is[5–7]:

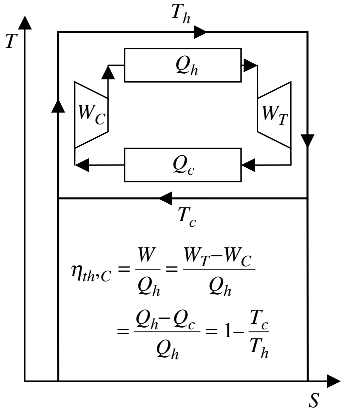


Fig. 12 Heat engine ideal Carnot cycle.

$$\begin{aligned}\underbrace{\frac{d}{dt} S_{\text{CV}}}_{\substack{\text{RATE OF ENTROPY}\\ \text{CHANGE IN CV}}} &= \underbrace{\sum_{\text{BS}} \frac{\dot{Q}_i}{T_i}}_{\substack{\text{BS TRANSFER}\\ \text{RATE WITH } Q}} + \underbrace{\sum_{\text{IN}} \dot{m}_j s_j}_{\substack{\text{TRANSPORT RATE}\\ \text{WITH MASS IN}}} \\ &\quad - \underbrace{\sum_{\text{OUT}} \dot{m}_k s_k}_{\substack{\text{TRANSPORT RATE}\\ \text{WITH MASS OUT}}} + \underbrace{\dot{S}_{\text{gen}}}_{\substack{\text{IRREVERSIBLE}\\ \text{GENERATION RATE}}}\end{aligned} \tag{21}$$

Heat Engines

Heat Engines are devices undergoing thermo-mechanical cycles (transforming thermal into mechanical energy), similar to one on Fig. 12, with mechanical expansion and compression net-work ($W{=}Q_h{-}Q_c$), obtained as the difference between the heat transferred to the engine from high temperature heat reservoir (at T_h) and rejected to a low (cold) temperature heat reservoir (at T_c), thus converting part of the thermal energy into mechanical work. In a close-system cycle the net-work-out is due to net-work of thermal-expansion and thermal-compression. Therefore, heat engine cycle cannot be accomplished without two thermal reservoirs at different temperatures, one at higher temperature to accomplish thermal expansion and work out, and another at lower temperature to accomplish thermal compression to initial volume and thus complete the cycle. The combustion process itself is an irreversible one, where chemical energy (electro-chemical energy binding atoms in reactants' molecules) is chaotically released during combustion, i.e., converted into random thermal energy of products' molecules, and cannot be fully converted into directional work energy. The *Second Law of Thermodynamics* limits the maximum amount of work that could be obtained from thermal energy between any two thermal reservoirs at different temperatures, hot T_h, and cold T_c, by using the ideal, reversible *Carnot cycle*, see Fig. 12, with thermal efficiency given by Eq. 22. As an example, consider $T_c = 293$ K and $T_h = 2273$ K:

$$\begin{aligned}\eta_{th,C\_ad} &= \frac{W}{Q_h} = \frac{Q_h - Q_c}{Q_h} \\ &= 1 - \frac{T_c}{T_h}\bigg|_{T_h = T_{ad} = 2273\ \text{K}, T_c = 293\ \text{K}} \\ &= 1 - \frac{293}{2273} = 87.1\% \end{aligned} \tag{22}$$

where, $W = W_T - W_C$, is the net-work of expansion, usually turbine (W_T), and compression (W_C). The maximum efficiency is achieved if heat is supplied at the highest possible temperature T_h, and released at the lowest possible temperature T_c. However, both temperatures are limited by the fact that fuel combustion is performed using oxygen with ambient air, resulting in maximum so called adiabatic, stoichiometric combustion temperature T_{ad}, which is for most fuels about 2000°C, or $T_{ad} = 2273$ K. A part of the heat supplied at hot temperature T_h, must be released to the surroundings at cold temperature about $T_c = 20°C = 293$ K, which results in a Carnot efficiency of 87.1%, see Eq. 22 and Fig. 12. However, the fuel heating value energy $Q_{HV} = Q_{ad\_var}$, is not all available at the adiabatic temperature of the products, but is distributed over their variable temperature range from initial surrounding temperature before combustion T_c, to final adiabatic temperature T_{ad}. For such a variable heat reservoir, a large number (infinite in the limit) of ideal Carnot engines operating at different temperatures (with $dW = dQ$), must be employed to achieve a reversible cycle, resulting in the variable-temperature Carnot cycle with the maximum possible combustion-products-to-work conversion efficiency[1]:

$$\begin{aligned}&\eta_{th,C\ var\ max} \\ &= \left(1 - \frac{\ln(T_{ad}/T_c)}{(T_{ad}/T_c) - 1}\right)\bigg|_{T_{ad} = 2273\ \text{K}, T_c = 293\ \text{K}} \\ &= 69.7\% \end{aligned} \tag{23}$$

The above Eq. 23 is valid if the cyclic medium has constant specific heat, otherwise integration will be required. Due to engine material property limitations and other unavoidable irreversibilities, it is impossible to reach the ideal Carnot efficiency. Different actual heat engines undergo similar but different cycles, depending on the system design. For example, internal combustion engines undergo the *Otto* cycle with gasoline fuel and the *Diesel* cycle with diesel fuel, while the steam and gas turbine power plants undergo the *Rankine* and *Brayton* cycles, respectively. However, with improvements in material properties, effective component cooling, and combining gas and steam turbine systems, the efficiencies above 50% are being achieved, which is a substantial improvement over the usual 30%–35% efficiency. The ideal Carnot cycle is an important reference to guide researchers and engineers to better understand limits and possibilities for new concepts and performance improvements of real heat engines.

CONCLUDING REMARKS: ENERGY PROVIDES EXISTENCE AND IS CAUSE FOR CHANGE

Energy is fundamental property of a physical system and refers to its potential to maintain a system identity or structure and to influence changes with other systems (via forced interaction) by imparting work (forced directional displacement) or heat (forced chaotic displacement/motion of a system molecular or related structures). Energy exists in many forms: electromagnetic (including light), electrical, magnetic, nuclear, chemical, thermal, and mechanical (including kinetic, elastic, gravitational, and sound); where, for example, electro-mechanical energy may be kinetic or potential, while thermal energy represents overall potential and chaotic motion energy of molecules and/or related micro structure.

The philosophical and practical aspects of energy and entropy, including reversibility and irreversibility could be summarized as follows:

- Energy is a fundamental concept indivisible from matter and space, and energy exchanges or transfers are associated with all processes (or changes), thus indivisible from time.
- Energy is "the building block" and fundamental property of matter and space, and thus a fundamental property of existence. For a given matter (system structure) and space (volume) energy defines the system equilibrium state, and vice versa.
- For a given system state (structure and phase) addition of energy will spontaneously tend to randomly redistribute (disorganize, degrade) over the system smaller microstructure and space it occupies, called thermal energy, equalizing the thermal energy–potential (temperature) and increasing the energy–displacement (entropy), and vice versa.
- Entropy may be transferred from system to system by reversible heat transfer and also generated due to irreversibility of heat and work transfer.

- Energy and mass are conserved within interacting systems (all of which may be considered as a combined isolated system not interacting with its surrounding systems), and energy transfer (in time) is irreversible (in one direction) from higher to lower energy–potential only, which then results in continuous generation (increase) of energy–displacement, called entropy generation, which is fundamental measure of irreversibility, or permanent changes.
- Reversible energy transfer is only possible as a limiting case of irreversible energy transfer at infinitesimally small energy–potential differences, thus in quasiequilibrium processes, with conservation of entropy. Since such changes are reversible, they are not permanent and, along with time, irrelevant.

In summary, energy is providing existence, and if exchanged, it has ability to perform change.

Glossary

Energy: It is fundamental property of a physical system and refers to its potential to maintain a system identity or structure and to influence changes with other systems (via forced-displacement interactions) by imparting work (forced directional displacement) or heat (forced chaotic displacement/motion of a system molecular or related structures). Energy exists in many forms: electromagnetic (including light), electrical, magnetic, nuclear, chemical, thermal, and mechanical (including kinetic, elastic, gravitational, and sound).

Energy Conservation: It may refer to the fundamental law of nature that energy and mass are conserved, i.e., cannot be created or destroyed but only transferred from one form or one system to another. Another meaning of energy conservation is improvement of efficiency of energy processes so that they could be accomplished with minimal use of energy sources and minimal impact on the environment.

Energy Conversion: A process of transformation of one form of energy to another, like conversion of chemical to thermal energy during combustion of fuels, or thermal to mechanical energy using the heat engines, etc.

Energy Efficiency: Ratio between useful (or minimally necessary) energy to complete a process and actual energy used to accomplish that process. Efficiency may also be defined as the ratio between energy used in an ideal energy-consuming process vs energy used in the corresponding real process, or vice versa for an energy-producing process. Energy, as per the conservation law, cannot be lost (destroyed), but part of energy input which is not converted into *useful energy* is customarily referred to as *energy loss*.

Entropy: It is an integral measure of thermal energy redistribution (due to heat transfer or irreversible heat generation) within a system mass and/or space (during system expansion), per absolute temperature level. Entropy is increasing from orderly crystalline structure at zero absolute temperature (zero reference) during reversible heating and entropy generation during irreversible energy conversion, i.e., energy degradation or random equi-partition within system material structure and space.

Exergy: It is the maximum system work potential if it is reversibly brought to the equilibrium with reference surroundings, i.e., exergy is a measure of a system non-equilibrium with regard to the reference system.

Heat: It is inevitable (spontaneous) energy transfer due to temperature differences (from higher to lower level), to a larger or smaller degree without control (dissipative) via chaotic (in all directions, non purposeful) displacement/motion of system molecules and related microstructure, as opposed to controlled (purposeful and directional) energy transfer referred to as *work* (see below).

Heat Engine: It is a device undergoing thermo-mechanical cycle that partially converts thermal energy into mechanical work and is limited by the *Carnot cycle* efficiency. The cycle mechanical expansion and compression net-work is obtained due to difference between heat transferred to the engine from a high temperature heat reservoir and rejected to a low temperature reservoir, thus converting part of thermal energy into mechanical work.

Mechanical Energy: It is defined as the energy associated with ordered motion of moving objects at large scale (kinetic) and ordered elastic potential energy within the material structure (potential elastic), as well as potential energy in gravitational field (potential gravitational).

Power: It is the energy rate per unit of time and is related to work or heat transfer processes (different work power or heating power).

System: (also *Particle* or *Body* or *Object*) refers to any, arbitrary chosen but fixed physical or material system in space (from a single particle to system of particles), which is subject to observation and analysis. System occupies so called system volume within its own enclosure interface or system boundary, and thus separates itself from its surroundings, i.e., other surrounding systems.

Temperature: It is a measure of the average quasi-translational kinetic energy associated with the disordered microscopic motion of atoms and molecules relevant for inter-particulate collision and heat transfer, thus temperature is related to relevant particle kinetic energy and not to the particle density in space.

Thermal Energy: It is defined as the energy associated with the random, disordered motion of molecules and potential energy due to intermolecular forces (also associated with phase change), as opposed to the macroscopic ordered energy associated with moving objects at large scale.

Total Internal Energy: It is defined as the energy associated with the random, disordered motion of molecules and intermolecular potential energy (thermal), "binding" potential energy associated with chemical molecular structure (chemical) and atomic nuclear structure (nuclear), as well as with other structural potentials in force fields (electrical, magnetic, etc.). It refers to the "invisible" microscopic energy on the subatomic, atomic and molecular scale.

Work: It is a type of controlled energy transfer when one system is exerting force in a specific direction and thus making a purposeful change (forced displacement) of the other systems. It is inevitably (spontaneously) accompanied, to a larger or smaller degree (negligible in ideal processes), with dissipative (without control) energy transfer referred to as *heat* (see above).

REFERENCES

1. Kostic, M. Work, power, and energy, In *Encyclopedia of Energy*, Cleveland, C.J., Ed.; Elsevier Inc.: Oxford, U.K., 2004; Vol. 6, 527–538.
2. Kostic, M. Irreversibility and reversible heat transfer: the quest and nature of energy and entropy, IMECE2004. In *ASME Proceedings*, ASME, New York, 2004.
3. Kostic, M. *Treatise with Reasoning Proof of the First Law of Energy Conservation*, Copyrighted Manuscript; Northern Illinois University: DeKalb, IL, 2006.
4. Kostic, M. *Treatise with Reasoning Proof of the Second Law of Energy Degradation*, Copyrighted Manuscript; Northern Illinois University: DeKalb, IL, 2006.
5. Moran, A.J.; Shapiro, H.N. *Fundamentals of Engineering Thermodynamics*, 4th Ed. Wiley: New York, 2000.
6. Bejan, A. *Advanced Engineering Thermodynamics*; Wiley: New York, 1988.
7. Cengel, Y.A.; Boles, M.A. *Thermodynamics, An Engineering Approach*, 5th Ed.; WCB McGraw-Hill: Boston, MA, 2006.
8. Zemansky, M.W.; Dittman, R.H. *Heat and Thermodynamics*; McGraw-Hill: New York, 1982.

Pricing Programs: Time-of-Use and Real Time

Ahmad Faruqui
The Brattle Group, San Francisco, California, U.S.A.

Abstract
This article surveys numerous pricing designs for improving economic efficiency in all market segments. Electricity is a very capital-intensive industry characterized by a significant peak load problem. Expensive generating plants have to be installed to meet peak loads that are only encountered for a few hundred hours a year. This raises the cost of electricity to all consumers. Average cost pricing, the staple of the industry in which rates do not vary by time of use, compounds the problem by creating cross-subsidies. Customers with flatter load shapes subsidize those with peakier load shapes.

The problem can be alleviated by modifying electricity pricing practices to allow time-variation in costs. This would provide customers an incentive to lower peak usage, either by curtailing or shifting their activities. In addition, it would eliminate unfair and economically unjustified cross subsidies. But the potential benefits of time-varying pricing have yet to be fully realized. Many barriers stand in the way of reform, including economic, technological and political. Of all these barriers, the most formidable ones are the political ones. They have to be resolved by modifying the legal and regulatory framework through which electricity pricing is determined.

INTRODUCTION

Time-of-use (TOU) pricing and real time pricing (RTP) programs are designed to lower system costs for utilities and bring down customer bills by raising prices during expensive hours and lowering them during inexpensive hours. They differ in, that the former fixes the price and time periods in advance while the latter fixes neither the price nor the time period in advance. Thus, TOU rates can be considered static while RTP rates can be considered dynamic, even though before feature time-varying prices. Other rate designs bridge the gap between these two rate designs, as shown below.

Time-of-use pricing (TOU). This rate design features prices that vary by time period, and are higher in peak periods and lower in off-peak periods. The simplest rate involves just two pricing seasons, with prices being higher during the peaking season. A time-of-day rate is slightly more complex and involves two pricing periods within a day, a peak period and an off-peak period. More complex rates have one or more shoulder periods and seasonal variation.

Critical peak pricing (CPP). This rate design layers a very high price during a few critical hours of the year. It can also be combined with a TOU rate. Typically, a CPP rate is only used on 12–15 days a year. These days are called the day before or the day of the critical peak price.

Extreme day pricing (EDP). This rate design is similar to CPP, except that the higher price is in effect for all 24 h for a maximum number of critical days, the timing of which is unknown until a day ahead.

Extreme day CPP (ED-CPP). This rate design is a variation of CPP in which the critical peak price applies to the critical peak hours on extreme days but there is no TOU pricing on other days.

Real time pricing (RTP). This rate design features prices that vary hourly or sub-hourly all year long, for some or all of a customer's load. Customers are notified of the rates on a day-ahead or hour-ahead basis.

Each of these rates exposes customers to varying amounts of price variance. Customers can lower their expected (average) price by taking more risks. For example, RTP rates are riskiest from the customer's viewpoint since they face wholesale prices that vary in real time, but they will most likely be associated with the lowest average price. Critical peak pricing rates carry less pricing uncertainty for customers, since customers know the prices ahead of time and the time for which these prices will be in effect is limited. However, the average price is likely to be higher than that for RTP rates. At the other end of the spectrum are rates that do not vary over the hours of the day and only vary seasonally. They provide the highest rate predictability to customers but are also likely to carry the highest average price.

TIME-OF-USE PRICING

Time-of-use pricing is commonplace in developed economies at all stages of market restructuring. Electricite de France (EDF) operates the most successful example of

Keywords: Electricity pricing; Rate design; Economic efficiency; Demand response.

TOU pricing. Currently, a third of its population of 30 million customers is estimated to be on TOU pricing. This pricing design was first introduced for residential customers in 1965 on a voluntary basis, having been first applied in the country to large industrial customers as the Green Tariff in 1956. The French model served for many years as a benchmark for many countries in Latin America. For example, in Brazil, it was introduced as the "Horo-sazonal" tariff, which divides the day into peak and off peak periods and the year into dry and wet seasons. The idea was to continue all the way to the residential customer (yellow tariff), but it never came to fruition.

Time-of-use rates have been mandatory in California for all customers above 500 kW since 1978, as a statewide policy response to the energy crisis of 1973. These rates are mandatory in several U.S. states but the size threshold varies by state.

Residential TOU rates are offered on a voluntary opt-in basis by utilities in all types of climates within the U.S., including Pepco in the Washington, DC area and the Salt River Project in the Phoenix area. The simplest variation involves two time periods. An example is the residential rate design offered by Pacific Gas & Electric Company (PG&E) in central and northern California. During the summer months, from noon to six P.M. on weekdays, electricity costs three times as much as during all other hours of the week. During the winter months, the price differential is smaller.

Another example is the project that was implemented by Puget Sound Energy (PSE) in the suburbs of Seattle. In May 2001, as a response to the power crisis in the Western states, PSE designed and implemented a TOU rate for its residential and small commercial customers. It involved four pricing periods. The morning and evening periods were the most expensive periods, followed by the mid-day period and the economy period. Unlike most TOU rates, which feature significant differentials between peak and off-peak prices, PSE's TOU rate featured very modest price differentials between the peak and off-peak periods, reflecting the hydro-based system in the Northwest.

The peak price was about 15% higher than the average price customers had faced prior to being moved to the TOU rate and the off-peak price was about 15% lower. To keep the rate simple, there was no seasonal variation in prices.

Puget Sound Energy placed about 300,000 customers on the rate, but they could opt-out to the standard rate if they so desired. There was no additional charge to participate in the rate. During the first year of the program, less than half of one percent elected to opt-out of the rate. Customer satisfaction with the rate was high. In focus groups, customers identified several benefits of the TOU rate besides bill savings, including greater control over their energy use; choice about which rate to be on; social responsibility; and energy security. PSE also provided a website to customers where they could review their load shapes for the past seven days.

Puget Sound Energy had a rate case settlement in June 2002. Under the terms of the settlement, the program became an opt-in program for new customers. The peak/off-peak rate differential of the TOU rate was reduced from 14 to 12 mils/kWh (A mil is a thousandth of a dollar). A monthly fee of $1 a month, about 80% of the estimated variable cost of providing TOU meter reading, was levied on participating customers. Finally, each quarter PSE would notify customers of their savings (or losses) on the program, and it would switch all customers to the lower-cost rate (flat or TOU) in August 2003.

In October 2002, PSE sent customers their first quarterly report. For 94% of the customers, this report showed that they were paying an extra 80 cents/month by participating in the TOU pilot, comprised of the difference between 20 cents of power cost savings and a dollar of incremental meter reading costs. This was marked in contrast to the first year of the program when, prior to charging customers any part of the TOU meter reading costs, over 55% of residential customers experienced bill savings by being on the TOU rate.

Even though the report was for a single quarter, 10% of the participating customers chose to opt-out of the program between July 1 and October 31. At the same time, 1.8% of new customers opted into the program.

Media coverage was very negative and featured interviews with customers claiming that they had shifted almost half of their load from peak to off-peak periods, only to find out that they had lost money. PSE pulled the plug on a program that had become the most visible national symbol of a utility's commitment to time-varying pricing, and agreed to refund the increased amounts to participating customers.

Lessons Learned From the PSE TOU Rate

Five lessons can be drawn from PSE's TOU program.

- Customers do shift loads in response to a TOU price signal, even if the price signal is quite modest. According to an independent analysis, customers consistently lowered peak period usage by 5% per month over a 15-month period.
- It is important to manage customer expectations about bill savings.
- Customers should be educated on the magnitude of bill savings they can expect from specific load shifting activities.
- It is desirable to conduct a pilot program involving a few thousand customers before offering a rate to hundreds of thousands of customers.
- Finally, and most importantly, any program should make a majority of the customers better off, or it should not be offered.

Developing a TOU Rate

It is fairly straightforward to develop a TOU rate design. The following sidebar shows the steps involved in developing a "revenue-neutral" TOU rate. Such a rate would leave the average customer's bill unchanged if that customer chose to make no adjustments in their pattern of usage. Of course, a customer who uses less power in the peak period than the average customer would be made better off (compared to his or her situation on the standard rate) by the rate even without responding to the rate and a customer who uses proportionately more power in the peak period than the average customer would be made worse off by the rate if he or she did not respond to the rate.

The sidebar brings out the type of information that is needed to develop a TOU rate.

CRITICAL PEAK PRICING

Under this rate design, customers are on TOU prices for most hours of the year but additionally face a much higher price during a small number of critical hours when system reliability is threatened or very high prices are encountered in wholesale markets because of extreme weather conditions and similar factors. In 1993, EDF (France) introduced a new rate design, tempo, and now has over 120,000 residential customers on it. The program features two daily pricing periods and three types of days. The year is divided into three types of days, named after the colors of the French flag. The blue days are the most numerous (300) and least expensive; the white days are the next most numerous (43) and mid-range in price; and the red days are the least numerous (22) and the most expensive. The ratio of prices between the most expensive time period (red peak hours) and the least expensive time period (blue off-peak hours) is about 15–1, reflecting the corresponding ratio in marginal costs.

The tempo rate does not offer a fixed calendar of days, but customers can learn what color will take effect the next day by checking a variety of different sources:

- Consulting the Tempo Internet website: www.tempo.tm.fr
- Subscribing to an email service that alerts them of the colors to come
- Using Minitel (a data terminal particular to France, sometimes called a primitive form of Internet)
- Using a vocal system over the telephone
- Checking an electrical device (*Compteur Electronique*) provided by EDF that can be plugged into any electrical socket.

The tempo rate was preceded by a pilot program, in which prices were quite a bit higher than those that were ultimately implemented. The rates associated with the tempo program and with EDF's standard TOU rate are shown in Fig. 1.

Sidebar 1 Developing a TOU rate involves several steps

Existing flat rate	
Per-customer revenue requirement	$100
Per-customer monthly usage	1000 kWh
Average price	$0.10/kWh
Revenue neutral TOU rate	
Estimate peak usage	200 kWh
Estimate off-peak usage	800 kWh
Set peak price = peak marginal cost	$0.20/kWh
Set off-peak price = off-peak marginal cost	$0.075 kWh
Given class revenue requirement	$100
Given monthly usage	1000 kWh
TOU rate with load shifting	
Estimate price elasticity	−0.2
Estimate new peak usage	160 kWh[a]
Estimate new off-peak usage	840 kWh
Estimate new monthly usage	1000 kWh
Estimate new per-customer monthly bill	$95
Estimate bill savings = per-customer revenue loss	$100 − $95 = $5

[a]These changes in usage for the peak and off-peak period are estimated by using the percent changes in peak and off-peak prices and the estimated price elasticity of demand.

Critical Peak Pricing With Enabling Technologies

Recently, a number of utilities have experimented with dynamic pricing options, sometimes in conjunction with enabling technologies that automate customer response during high priced periods. As seen below, dynamic pricing, especially when combined with enabling technologies, can produce much larger reductions in peak demand than traditional TOU or non-technology enabled CPP rates.

Two utilities, GPU in Pennsylvania and American Electric Power in Ohio, conducted small-scale pilot programs in the 1980s using a two-way communication and control technology called TransText. The TransText device allows for the creation of a fourth critical price period in which the retail price of electricity rises to a much higher level (e.g., 50 cents/kWh in the GPU pilot). The number of hours during which this price can be charged is small (e.g., 100–200 h) and the customer knows what the critical price will be ahead of time, but does not know when the price may be called.

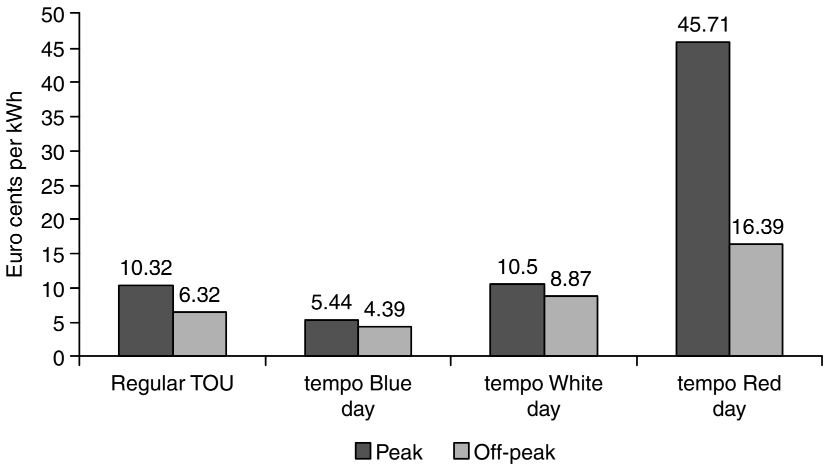


Fig. 1 EDF's *tempo* and standard TOU rates.

The TransText device incorporates an advanced communication feature that lets customers know that a critical period is approaching and it can be programmed so that the customer's thermostat is automatically adjusted when prices exceed a certain level. Using this technology, American Electric Power found significant load shifting, with estimated peak demand reductions of 2–3 kW per customer during on-peak periods and of 3.5–6.6 kW during critical peak periods. These critical peak reductions represented a drop of nearly 60% of a typical customer's peak load during the winter period.

The GPU experiment produced similar results, showing elasticities of substitution that ranged from −0.31 to −0.40, significantly higher than the elasticities associated with traditional TOU rates, which have averaged −0.14 in a range of studies. These elasticities were estimated by comparing customer loads on days when control was being exercised with days when control was not being exercised.

Another example is provided by Gulf Power Company's *Good Cents Select* program in Florida. Like the GPU experiment, the Gulf Power program uses dynamic pricing to obtain additional benefits beyond traditional TOU pricing. Under this voluntary program, residential consumers face a three-part TOU rate for 99% of all hours in the year, where the peak period price of $0.093/kWh is roughly 60% higher than the standard (flat) tariff price and approximately twice the intermediate (shoulder) price. For the remaining 1% of the hours, Gulf Power has the option of charging a critical period price equal to $0.29/kWh, more than three times the value of the peak-period price. The timing of this much higher price is uncertain and it is called during the day when critical conditions are encountered. In conjunction with this rate, participating customers are provided with a programmable/controllable thermostat that automatically adjusts their heating and cooling loads and up to three additional control points in the home such as water heating and pool pumps. The devices can be programmed to modify usage when prices exceed a certain level.

Gulf Power is seeing results similar to those of the GPU experiment. Peak-period reductions in energy use over a 2-year period have equaled roughly 22% compared with a control group, while reductions during critical-peak periods have equaled almost 42%. Diversified coincident peak demand reductions have exceeded more than 2 kW per customer. This voluntary program has been in place for less than a year, and Gulf Power has already signed up more than 3000 high use customers. It hopes to attract 40,000 customers over the next 10 years, representing about 10% of the residential population. Participating customers pay roughly $5/month to help offset the additional cost of the communication and control equipment. In a recent survey, the program received a 96% satisfaction rating.

The Gulf Power program is targeted at high use customers, just like the EDF program. Customer savings are large enough to offset the program costs. Both rates have significant peak to off-peak differentials as well. Because of these two factors, the programs have been successful. The PSE program failed in part because it had weak peak to off-peak differential and in part because it did not target the large customers.

California's Pricing Experiment

The state of California conducted a statewide pricing pilot (SPP) during the 2003–2005 timeframe to test customer response to a variety of pricing options, including TOU rates and CPP rates. In California, standard residential tariffs involve an "inverted tier" design in which the price of power rises with electricity usage. The typical residential customer pays an average price of about 13 cents/kWh. Within the SPP, customers on TOU and CPP rates pay a higher price during the five-hour peak period that lasts from 2 P.M. to 7 P.M. on weekdays and a lower price during the off-peak period, which applies during all other hours.

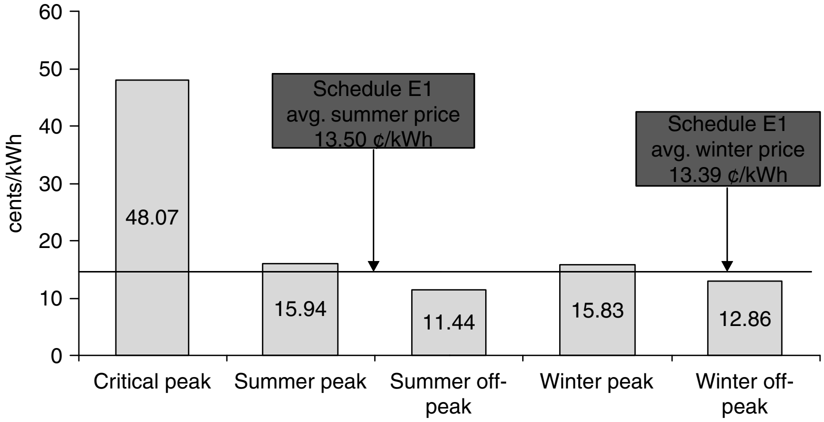


Fig. 2 Critical-peak pricing (CPP) tariff.

Each TOU and CPP rate involves two sets of peak/off-peak prices, to allow for precise estimation of the elasticities of demand. On average, customers on TOU rates are given a discount of 23% during the off-peak hours and are charged a price of around 10 cents. They are charged a price of 22 cents during the peak hours, which is 69% higher than their standard rate. Thus, with TOU rates, customers are given a strong incentive to curtail peak usage and to shift usage to off-peak periods. However, the incentive is much greater on selected days for customers on CPP rates, who are charged, on average, a price of 64 cents during the peak hours on 12 summer days, i.e., prices are nearly five times higher than the standard price. On the peak hours of other days and the off-peak hours of all days they face prices that are slightly lower than the prices faced by TOU customers during these periods. Fig. 2 shows the CPP tariffs that were used in the California experiment.

Analysis of data from the California experiment indicates that CPP rate customers face “rifle shot” price signals that can be very effective at reducing peak demand, thus dampening wholesale prices and obviating the need for building costly power plants that would run for only a few hundred hours a year. Customers are likely to respond to higher peak prices by reducing peak usage, e.g., by reducing air conditioning usage, and perhaps by shifting some peak period usage associated with laundry, dish-washing and cooking activities to lower cost off-peak periods. They may also be raising off-peak use in response to lower off-peak rates by raising air conditioning usage, increasing lighting levels, and so on. Finally, since prices have changed in the peak and off-peak periods, the average price for electricity over the day may have changed for some customers as well. This would trigger additional changes in usage.

Fig. 3 shows the changes in customer load shapes caused by the CPP tariff in customers who were located in the San Diego Gas and Electric service area. The black line shows the usage of the control group of customers. The gray line shows the usage of customers who were equipped with a smart thermostat that received a communication signal from the utility during critical hours, which raised the set point of the thermostat. Their tariff was unchanged from that of the control group. The difference between the two lines is noticeable and suggests that remotely controlling the thermostat lowers peak usage. The white line shows the

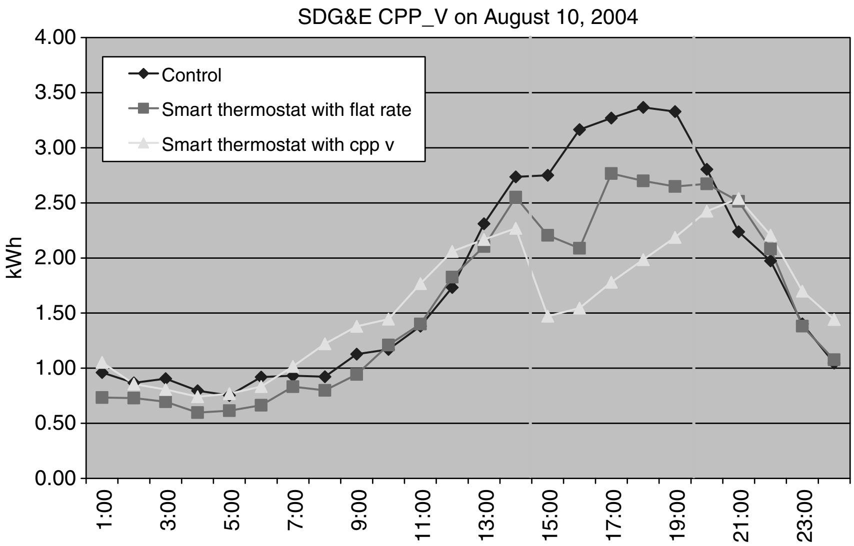


Fig. 3 Changes in customer load shapes.

usage of customers who were equipped with a smart thermostat and who also were placed on the CPP tariff. They show a greater drop than customers who had the smart thermostat but who were not placed on the CPP tariff.

REAL-TIME PRICING FOR RESIDENTIAL CUSTOMERS

The Chicago Community Energy Cooperative (Co-op) has implemented a market-based RTP pricing plan for residential customers, in conjunction with the local electric utility, CommEd. The utility provides the rate and the metering/billing system while the Co-op provides customer notification (via a web site, e-mail and telephone), education, and energy management tools (Fig. 4).

The pilot program is intended to model the bundled rate/market rate differential in the post-2006 market environment when the rate freeze is lifted. It involves RTP prices on a day-ahead basis for the generation portion of the rate. The prices are capped at 50 cents/kWh. The project is designed to estimate the magnitude of customer response to hourly energy pricing and understand the drivers of responsiveness. This is a 3-year experimental program that commenced in January 2003.

In the first year of the program, 750 customers were enrolled. Of these, 100 are in a control group. The summer of 2003 was mild in terms of both temperatures and prices. For example, the number of days with a maximum temperature higher than 90°C was 10 versus a historical average of 18. The maximum price was 12.39 cents/kWh, versus a price of 38.11 cents/kWh during the crisis years of 2000–2002.

Analysis of customer loads during the first year indicates that participants responded to the higher prices they faced during the peak periods. A price elasticity of −0.042 was estimated over the full range of prices. Over half of all participants showed significant response to high price notifications. Aggregate demand reduction was as high as 25% during the notification period. Over 80% of the participants reported modifying their air conditioning usage, and over 70% reported modifying their clothes-washing patterns.

Multifamily households as a group were more price-responsive than single-family households. Households with window air conditioners maintained their price responsiveness better across multiple high-priced hours than single-family households, who started out strong but whose responsiveness tended to taper off during the high priced periods.

Customer satisfaction was very high with the program. The program was “quick and easy” for 82% of the participants and “time consuming and difficult” for 1%. Participants saved on average $12/month or 20% of their monthly bill.

The project has shown that residential customers are a viable market for RTP. They represent a key target market, since residential load is a major contributor to system peak. And giving residential customers a choice of pricing options may be the only way to give them a meaningful choice in restructured power markets.

REAL-TIME PRICING FOR COMMERCIAL AND INDUSTRIAL CUSTOMERS

Utilities in the Southeastern U.S. have implemented RTP rates for about 2000 customers on a day-ahead or hour-ahead basis. These companies include Georgia Power,

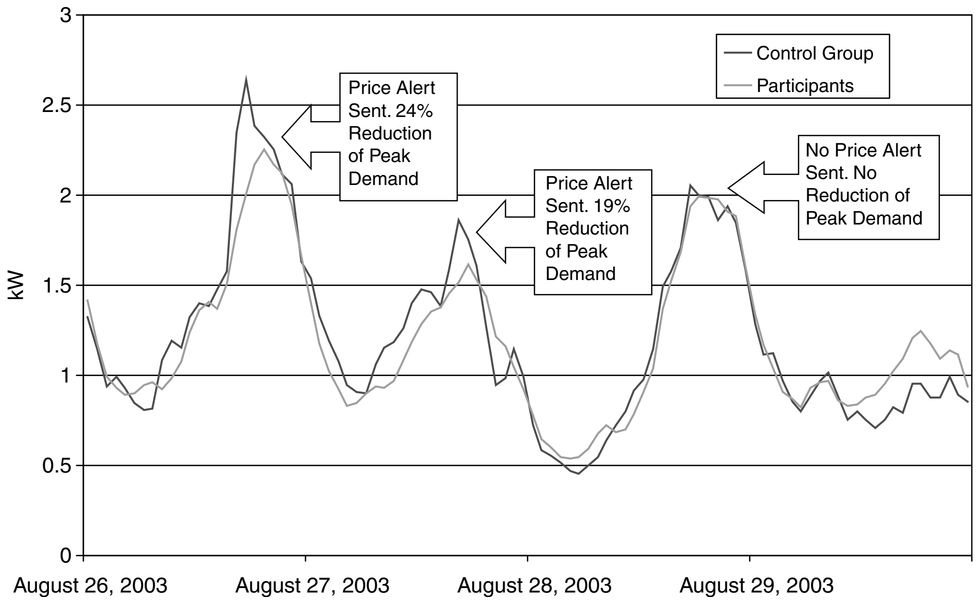


Fig. 4 Impact of real time pricing on Chicago Residences.

Duke Power and the Tennessee Valley Authority. The Georgia Power program is discussed in detail below.

Before describing the Georgia Power program, we note that RTP rates were probably first used by ESKOM, the state-owned utility in South Africa, for its largest customers, including the fabled gold mines. ESKOM has 1400 MW of load on day-ahead RTP. These customers drop their load by 350–400 MW for up to three consecutive hours when faced with high prices. While RTP is set up on a day-ahead basis, customer response is not used to optimize the dispatch of the power system. Electricity prices are based on the Pool Output Price, and do not change in response to changes in customer demand that may be induced by RTP. The utility is not aggressively marketing the program for this reason. It hopes that once a competitive energy market has been created, with a functioning Power Exchange, RTP will then be able to play its proper role in system operations.

If RTP had been implemented in California during the summer of 2000, much of the power crisis that developed in May 2000 would have abated within a month, rather than persisting for a year. If only a small proportion of the total customers had bought power on RTP, statewide peak demand would have dropped by 2.5%, or 1250 MW. During the peak hours, this would have lowered wholesale market prices by 20%. The state's power costs for the summer would have dropped by 6% (Faruqui, Ahmad, Hung-po Chao, Vic Niemeyer, Jeremy Platt and Karl Stahlkopf, "Getting out of the Dark," *Regulation*, Fall 2001, pp. 58–62).

RTP at Georgia Power

Georgia Power runs the worlds largest and possibly the most successful RTP program. The company estimates that during emergency conditions, its customers drop demand by 17%, freeing up 800 MW of capacity. A load drop of this magnitude eliminates the need for several expensive power plants that would otherwise be needed for meeting the peak load.

Program Background

Georgia law permits customers with 900 kW or more of connected load to put their load out to bid, and be served by any supplier in the state. In the late 1980s, Georgia Power was competing for these customers with almost 100 rural cooperatives and municipal utilities. In part to increase its competitiveness, Georgia Power began looking into RTP. In 1992, it began a 2-year controlled pilot, with the goals of increasing competitiveness; improving customer satisfaction by giving customers more control over their bills; and curtailing load when needed to balance supply and demand.

Georgia Power was one of the first utilities in the country to develop a two-part RTP tariff, following the lead of Niagara Mohawk in New York that had launched an Hourly Pricing Program in the late 1980s. The utility chose a two-part rather than a one-part rate for several reasons. First, the two-part rate allows the hourly price to more closely reflect the utility's true marginal cost. Second, the two-part rate best represents the "market price." Georgia Power believed a two-part rate would give it an opportunity to work with customers on price protection products. A discussion of price protection products is provided below. In addition, the utility was concerned about revenue stability; with a one-part rate, it would lose some of the contribution to fixed costs when customers curtailed in high priced hours. Georgia Power has expanded its RTP offerings since the 1992 pilot, but the basics of the program and tariff have remained relatively unchanged for almost a decade.

Rate Structure

Customers are billed for "baseline" use at their standard rate and pay (or receive credits) for energy used above (or below) the baseline each hour at the hourly price. The hourly price is composed of a measure of marginal energy costs, line losses, a "risk recovery factor" for forecasting risk (a fixed adder), and—near peaks—marginal transmission costs and outage cost estimates. Marginal transmission costs are triggered by load and temperature. Outage cost estimates are based on loss of load probabilities, as well as customer surveys on the costs of having an outage.

Georgia Power offers a "day-ahead" program, where customers are notified of price schedules by 4 P.M. the day before they go into effect, and an "hour-ahead" program, where customers are given an hour's notice on price. Currently, interruptible customers are served on the hour-ahead program. For these customers, their customer baseline (CBL) drops to their firm contract level during periods of interruption. Customers who do not interrupt to their firm levels pay interruption penalties plus the hourly prices. The utility has filed a tariff with the Public Service Commission that would allow interruptible customers on the day-ahead rate as well. The other difference between the day and hour-ahead rates is that the risk-recovery factor for the day-ahead rate is greater than that for the hour-ahead rate (4 mils/kWh versus 3 mils/kWh), since the utility bears a greater forecast risk.

Setting the Customer Baseline

When Georgia Power began its RTP program, it based a customer's baseline usage, or CBL, on an 8760-point hourly load profile. However, customers often found this CBL confusing, and therefore frustrating. In response to these customers, Georgia Power now offers 360-point CBLs (with 24 average hourly weekday loads per month and six average four-hour weekend day loads, for a total of 30 CBL points per month), and two-point CBLs.

The two-point CBLs simply average usage levels during the peak and off-peak periods.

The majority of customers (basically, the high-load-factor customers) now select the two-point CBL. If the two-point CBL does not seem appropriate based on a customer's usage profile, Georgia Power will usually use a 360-point CBL. Only a very few "unique loads" use the 8760-point CBL today (Our source noted that customers who can "really respond a lot" are typically on the higher point CBLs).

Price Protection Products

Georgia Power offers customers a variety of products that allow customers to influence their exposure to RTP price risk. One product, the adjustable CBL, allows customers to temporarily adjust their CBLs. For example, if customers want to lower their exposure to price volatility, they would increase CBLs. (Customers wanting to raise their CBLs must be on the RTP rate for a year, so that Georgia Power can determine how high the CBL can be raised.) Customers wanting to expose more loads to real-time prices—presumably because they believe it will be a cool summer—can lower their CBLs. Of the roughly 1650 customers on RTP, 600 currently have adjustable CBLs. About 60% of the incremental energy sold on the RTP rate (i.e., usage above baseline) is now protected by this product.

Georgia Power also offers a variety of financial products to limit customers' exposure to RTP price volatility. These products include price caps, contracts for differences, collars, index swaps, and index caps (Georgia Power's price-cap product guarantees that average RTP prices over a specific time period will not go above the cap. Its contract for differences gives a fixed price guarantee on the average RTP price. The collar has a cap and floor on the average RTP price over a specific time period. The index swap is a financial agreement that ties the RTP price to a commodity price index. If the commodity price index increases, so does the RTP price. If it decreases, so does the RTP price. The index cap is a financial agreement that ties an RTP price cap to a commodity price index. As the commodity price increases or decreases, so does the price cap). Georgia Power has sold these Price Protection Products, or PPPs, for 3 years. It currently has 250 contracts with about 90 customers. (Customers have multiple contracts to cover different time periods.) Georgia Power believes that offering these products has not probably increased the number of customers on the RTP program, but it has increased customer satisfaction. The utility has examined whether offering the PPPs has dampened price responsiveness, and has found no evidence of this.

LESSONS LEARNED

Georgia Power's experience highlights a number of lessons that have also been seen at other utilities. First, RTP can deliver substantial peak savingsssas, despite the fact that many customers are not very responsive to price. When the hourly price reached $6.40/kWh, Georgia Power saw 850 MW of load reduction (out of 1500–2000 MW of incremental or above-baseline load) from its RTP customers. Georgia Power also believes that customers have responded to the availability of low off-peak prices by expanding their facilities and business operations in Georgia. In other words, the rate has served to bring economic growth to the state and been a form of strategic electrification while also being a form of load management.

The utility's experience also supports the finding that customers join RTP programs to have access to lower cost power. When hourly prices went up in response to changing market conditions, customers sought price relief, and were granted it by the Georgia Public Services Commission.

Georgia Power has also found that a small percentage of customers are willing to pay for limited protection against price volatility. In response to customer requests, they developed and now sell a variety of risk-management products.

Georgia Power has also found that manufacturers with highly energy-intensive processes, such as chemical and pulp and paper companies are generally the most price responsive customers. It is also learnt that some commercial customers would respond to price. Office buildings, universities, grocery stores, and even a hospital (that changes chiller use based on hourly prices) are all responsive to real-time pricing.

Georgia Power states that the major lesson it has learnt is that education is the key to a successful RTP program. Georgia Power now holds annual, statewide meetings with all its customers to keep them informed about the RTP program. The utility believes its education program has paid off in customer satisfaction.

CONCLUSION

Electricity is a very capital-intensive industry characterized by a significant peak load problem. Expensive generating plants have to be installed to meet peak loads that are only encountered for a few hundred hours a year. This raises the cost of electricity to all consumers. Average cost pricing, the staple of the industry in which rates do not vary by time of use, compounds the problem by creating cross-subsidies. Customers with flatter load shapes subsidize those with peakier load shapes.

The problem can be alleviated by modifying electricity pricing practices to allow time-variation in costs. This would provide customers an incentive to lower peak usage, either by curtailing or shifting their activities. In addition, it would eliminate unfair and economically unjustified cross subsidies. As surveyed in this article, there are numerous pricing designs for improving economic efficiency in all market segments. But the potential benefits of time-varying pricing have yet to be fully realized. Many barriers stand in the way of reform, including economic, technological and political. Of all these barriers, the most formidable ones are the political ones. They have to be resolved by modifying the legal and regulatory framework through which electricity pricing is determined.

ACKNOWLEDGMENTS

Principal, The Brattle Group. This paper has benefited from conversations over the years with several colleagues. Without implicating them for any errors, I would like to specifically acknowledge the contributions of Andrew Bell, Steve Braithwait, Kelly Eakin, Robert Earle, Stephen S. George, Richard Goldberg, Phil Hanser, Roger Levy, Larry Ruff and Lisa Wood.

BIBLIOGRAPHY

1. Braithwait, S.; Faruqui, A. The choice not to buy: Energy savings and policy alternatives for demand response. Public Utilities Fortnightly **2001**, *March 15*.
2. Earle, R.; Faruqui, A. Toward a new paradigm for valuing demand response. The Electricity Journal **2006**, *19* (4), 21–31.
3. Faruqui, A.; Eakin, K.; Eds. *Electricity Pricing in Transition*, Kluwer Academic Publishing: Amsterdam, U.K., 2002.
4. Faruqui, A.; Eakin, K.; Eds. *Pricing in Competitive Electricity Markets*, Kluwer Academic Publishing: Amsterdam, U.K., 2000.
5. Faruqui, A.; George, S.S. Pushing the envelope on rate design. The Electricity Journal **2006**, *19* (2), 33–42.
6. Faruqui, A.; George, S.S. Quantifying customer response to dynamic pricing. The Electricity Journal **2005**, *18* (4), 53–63.
7. Faruqui, A.; George, S.S. Dynamic pricing for the mass market: California experiment. Public Utilities Fortnightly **2003**, *141* (13), 33–35.
8. Faruqui, A. Toward post-modern pricing. The Electricity Journal **2003**, *16* (6) [Guest Editorial].
9. Faruqui, A.; George, S.S. Demise of PSE's TOU program imparts lessons. Electric Light and Power **2003**, *January*.
10. Faruqui, A.; George, S.S. The value of dynamic pricing. The Electricity Journal **2002**, *15* (6).
11. Fox-Penner, P.; Broehm, R. Price-responsive electric demand: A national necessity, not an option. In *Electricity Pricing in Transition*; Faruqui, A., Eakin, B.K., Eds., Kluwer Academic Publishers: New York, 2002; [Chapter 10].
12. Fox-Penner, P.; Hanser, P.Q.; Wharton, J.B. Real-Time Pricing: Restructuring's Big Bang? Public Utilities Fortnightly March 1997.
13. Hanser, P.Q.; Caves, D.W.; Herriges, J.; Windle, R.J. Load impact of interruptible and curtailable rate programs. IEEE Transactions on Power Systems **1988**, *3* (4).
14. Wood, L. The new vanilla: Why making time-of-use the default rate for residential customers makes sense. Energy Customer Management **2002**, *5* (1).
15. Wood, L. *Effective Demand Response Programs for Mass Market Customers*. Presented at NYSERDA Time Sensitive Electricity Pricing Workshop, Albany, NY, October 3, 2002.
16. Wood, L. *The Customer's Critical Role in Demand Response*, Presented at U.S. Association for Energy Economics' (National Capital Area Chapter) New Developments in Electric Market Restructuring Conference, Washington, DC, May 21, 2002.

Psychrometrics

S. A. Sherif
Department of Mechanical and Aerospace Engineering, Wayne K. and Lyla L. Masur HVAC Laboratory, University of Florida, Gainesville, Florida, U.S.A.

Abstract
This entry presents the basics of psychrometric theory. This includes a brief discussion of moist air properties and psychrometric processes as well as provides examples of how the theory may be used in the HVAC design process for different summer and winter systems.

INTRODUCTION

Psychrometrics is the science of moist air properties and processes. It is used to illustrate and analyze air conditioning cycles, translating the knowledge of the building heating or cooling loads (which are in kW or tons) into volume flow rates (in m^3/s or cfm) for the air to be circulated into the duct system. The approximate composition of dry air by volume is as follows: nitrogen = 79.08%; oxygen = 20.95%; argon = 0.93%; carbon dioxide = 0.03%; other gases = 0.01%. Water vapor is lighter than dry air. The amount of water vapor that the air can carry increases with its temperature. Any amount of moisture that is present beyond what the air can carry at the prevailing temperature can only exist in the liquid phase as suspended liquid droplets (if the air temperature is above the freezing point of water) or in the solid state as suspended ice crystals (if the temperature is below the freezing point).

The most exact formulations of thermodynamic properties of moist air in the temperature range of −100°C–200°C are based on the study performed by Hyland and Wexler.[4,5] More recent studies by Sauer et al.[8] and Nelson et al.[7] provided psychrometric data for moist air in the temperature range 200°C–320°C and humidity ratios from 0 to 1 kg/kg_{air} at pressures of 0.07706 MPa (corresponding to an altitude of 2250 m), 0.101325, 0.2, 1.0, and 5 MPa. Both studies developed the psychrometric data using the most current values of the virial coefficients, enthalpy, and entropy of both air and water vapor. Other psychrometric data were generated by Stewart et al.,[9] who created psychrometric charts in SI units at low pressures. More psychrometric charts and tables are also available in the American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE) brochure on psychrometry.[1] Most recently, Mago and Sherif[6] presented new psychrometric charts and property formulations of supersaturated air in the temperature range from −40 to +40°C.

The most commonly used psychrometric quantities include the dry- and wet-bulb temperatures, dew point, humidity ratio, absolute humidity, specific humidity, relative humidity, and degree of saturation. These will be briefly defined and discussed.

Dry-Bulb Temperature

This is the temperature measured by a dry-bulb thermometer. There are several temperature scales commonly used in measuring the temperature. In the inch–pound (I–P) system of units, at standard atmosphere, the Fahrenheit scale has a water freezing point of 32°F and a boiling point of 212°F. In the International System (SI) of units, the Celsius scale has a water freezing point of 0°C and a boiling point of 100°C.

Wet-Bulb Temperature and Thermodynamic Wet-Bulb Temperature

The wet-bulb temperature is the temperature measured with a wet-bulb thermometer after the reading has stabilized in the air stream. Because of the evaporative cooling effect, the temperature measured with a wet-bulb thermometer is lower than the dry-bulb temperature, except when the air is saturated. In that case, the wet-bulb and dry-bulb temperatures are the same. The thermodynamic wet-bulb temperature, on the other hand, is the saturation temperature of moist air at the end of an ideal adiabatic saturation process. The latter process is defined as one of saturating an air stream by passing it over a water surface of infinite length in a well-insulated chamber.

Dew-Point Temperature

This is the temperature at which moisture will begin to condense out of the air.

Keywords: Psychrometrics; Moist air; Air conditioning; Psychrometric processes.

Saturation Pressure

Saturation pressure is needed to determine a number of moist air properties. Between the triple-point and critical-point temperatures of water, two states (saturated liquid and saturated vapor) may coexist in equilibrium. The saturation pressure over ice for the temperature range $-100°C \le T \le 0°C$ is given by[2]

$$\log_e(P_s) = \frac{a_0}{T + 273.15} + a_1 + a_2(T + 273.15) + a_3(T + 273.15)^2 + a_4(T + 273.15)^3 + a_5(T + 273.15)^4 + a_6\log_e(T + 273.15)$$

where $a_0 = -0.56745359 \times 10^4$; $a_1 = 6.3925247$; $a_2 = -0.9677843 \times 10^{-2}$; $a_3 = 0.62215701 \times 10^{-6}$; $a_4 = 0.20747825 \times 10^{-8}$; $a_5 = -0.9484024 \times 10^{-12}$; $a_6 = 4.1635019$.

The saturation pressure over liquid water for the temperature range of $0°C \le T \le 200°C$ can be written as[2]

$$\log_e(P_s) = \frac{b_0}{T + 273.15} + b_1 + b_2(T + 273.15) + b_3(T + 273.15)^2 + b_4(T + 273.15)^3 + b_5\log_e(T + 273.15)$$

where $b_0 = -0.58002206 \times 10^4$; $b_1 = 1.3914993$; $b_2 = -4.8640239 \times 10^{-2}$; $b_3 = 0.41764768 \times 10^{-4}$; $b_4 = -0.14452093 \times 10^{-7}$; $b_5 = 6.5459673$.

In both of the above equations, P_s is in Pa and T is in °C.

Humidity Ratio, Specific Humidity, Absolute Humidity, and Relative Humidity

The humidity ratio is the ratio of the mass of water vapor to the mass of dry air contained in the mixture of moist air. This is expressed as follows:

$$W_m = \frac{m_w}{m_a} = 0.62198\,\frac{\varepsilon(P_s\phi)}{P - \varepsilon(P_s\phi)}$$

where m_w and m_a are the mass of water vapor and dry air, respectively, P_s is the saturation pressure of the water vapor, P is the total pressure, ϕ is the relative humidity of the air, and ε is an enhancement factor. The enhancement factor (ε) is a correction parameter that takes into account the effect of dissolved gases and pressure on the properties of the condensed phase as well as the effect of intermolecular forces on the properties of moisture itself.[2] The enhancement factor can be expressed in terms of virial coefficients, but the equation is rather complicated. An equally accurate but simpler expression can be determined from a polynomial equation which has been least-square curve-fitted to the data, as shown below.[6]

$$\varepsilon = 1.00391 - 7.82205 \times 10^{-6}T + 6.94682 \times 10^{-7}T^2 + 3.04059 \times 10^{-9}T^3 - 2.7852 \times 10^{-11}T^4 - 5.656 \times 10^{-13}T^5$$

where T as given above is the dry-bulb temperature in °C.

The specific humidity is the ratio of the mass of water vapor to the total mass of the moist air sample. This is expressed as follows:

$$\gamma = \frac{m_w}{(m_w + m_a)}$$

The absolute humidity is the ratio of the mass of water vapor to the total volume of the sample. This is expressed as follows:

$$d_v = \frac{m_w}{V}$$

The relative humidity is the ratio of the mole fraction of water vapor in a moist air sample to the mole fraction of water vapor in a saturated moist air sample at the same temperature and pressure. According to ASHRAE,[2] this is expressed as follows:

$$\phi = \left(\frac{x_w}{x_{ws}}\right)_{T,P}$$

where the subscript s in the denominator refers to the saturation condition.

Degree of Saturation

This is the ratio of the humidity ratio of moist air to the humidity ratio of saturated moist air at the same temperature and pressure. This is expressed as follows:

$$\mu_d = \frac{W_m}{W_s}$$

where W_m is the humidity ratio of moist air while W_s is the humidity ratio of saturated air at the same temperature and pressure. The degree of saturation may have a value between 0 and 1 for subsaturated air. Combining the definition of the degree of saturation and those of the humidity parameters above, the following equation can be reached:

$$\mu_d = \frac{\phi}{1 + (1 - \phi)W_s/0.62198}$$

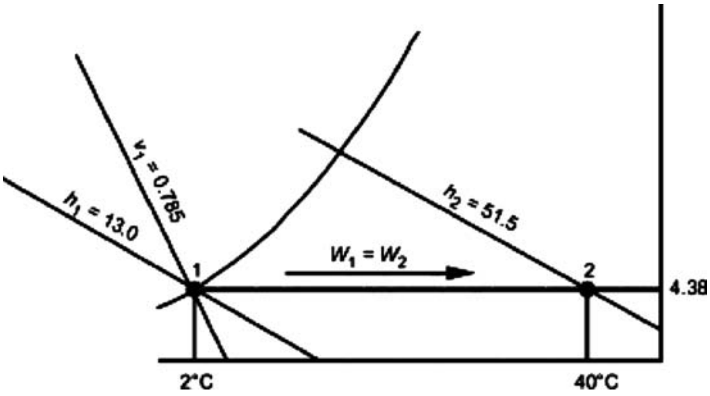


Fig. 1 Psychrometric depiction of sensible heating (1–2) and cooling (2–1).

PSYCHROMETRIC PROCESSES

Sensible Heating or Cooling

This is the process of heating or cooling the air without changing its moisture content. It is represented by lines of constant humidity ratio on the psychrometric chart (see Fig. 1). Sensible heating is accomplished when the air passes over a heating coil. Sensible cooling is accomplished when the air passes over a cooling coil whose surface temperature is above the dew point temperature of the air.

Humidification (with Heating or Cooling)

This is the process of introducing moisture into the air stream. In the winter season, humidification is frequently required because the cold outside air, infiltrating into a heated space or being intentionally brought in to satisfy the space ventilation requirement, is too dry. In the summer season, on the other hand, humidification is usually done as part of an evaporative cooling system. Humidification is achieved in a variety of ways, which range from using spray washers to passing the air over a pool of water to injecting steam. The process is represented on the psychrometric chart as a line of constant wet-bulb temperature when the water being sprayed is not externally heated or cooled. When there is external heating or cooling, the process is represented by a line to the right or to the left of the wet-bulb temperature line, respectively. Depending on the magnitude of water heating, the humidification-process line can be oriented in such a way that results in an increase in the dry-bulb temperature of the air stream at the exit of the dehumidifying device. In the extreme case of spraying cold water in the air stream such that the water temperature is less than the dew point temperature of water vapor in air, the spraying process will actually result in air dehumidification.

Cooling and Dehumidification

This process is used in air conditioning systems operating in hot and humid climates. It is accomplished by using a cooling coil whose surface temperature is below the dew point temperature of water vapor in air. On the psychrometric chart, the process is represented by a line in which both the dry-bulb temperature and the humidity ratio decrease (see Fig. 2). Because of the fact that not all of the air molecules going through the cooling coil make physical contact with the coil surface, the air condition at the exit of the coil is usually not saturated (but close to the saturation curve). This fact is reflected by the use of the so-called coil bypass factor, which is defined as the ratio of the temperature difference between the leaving coil air condition and the coil apparatus dew point and that between the entering air condition and coil apparatus dew point (see Fig. 3). This is expressed as follows:

$$B_F = \frac{T_{cc} - T_{cadp}}{T_m - T_{cadp}}$$

Coil bypass factors therefore range from 0 to 1. Fig. 3 shows the apparatus dew point and the coil bypass factor.

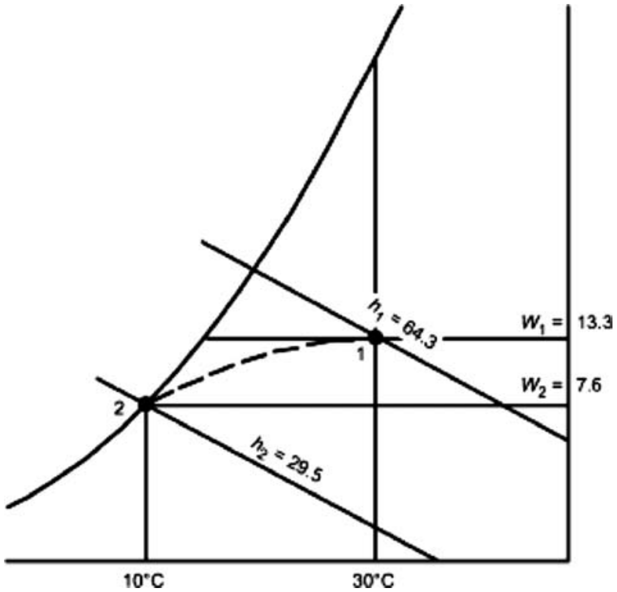


Fig. 2 Psychrometric depiction of the cooling and dehumidifying process (1–2).

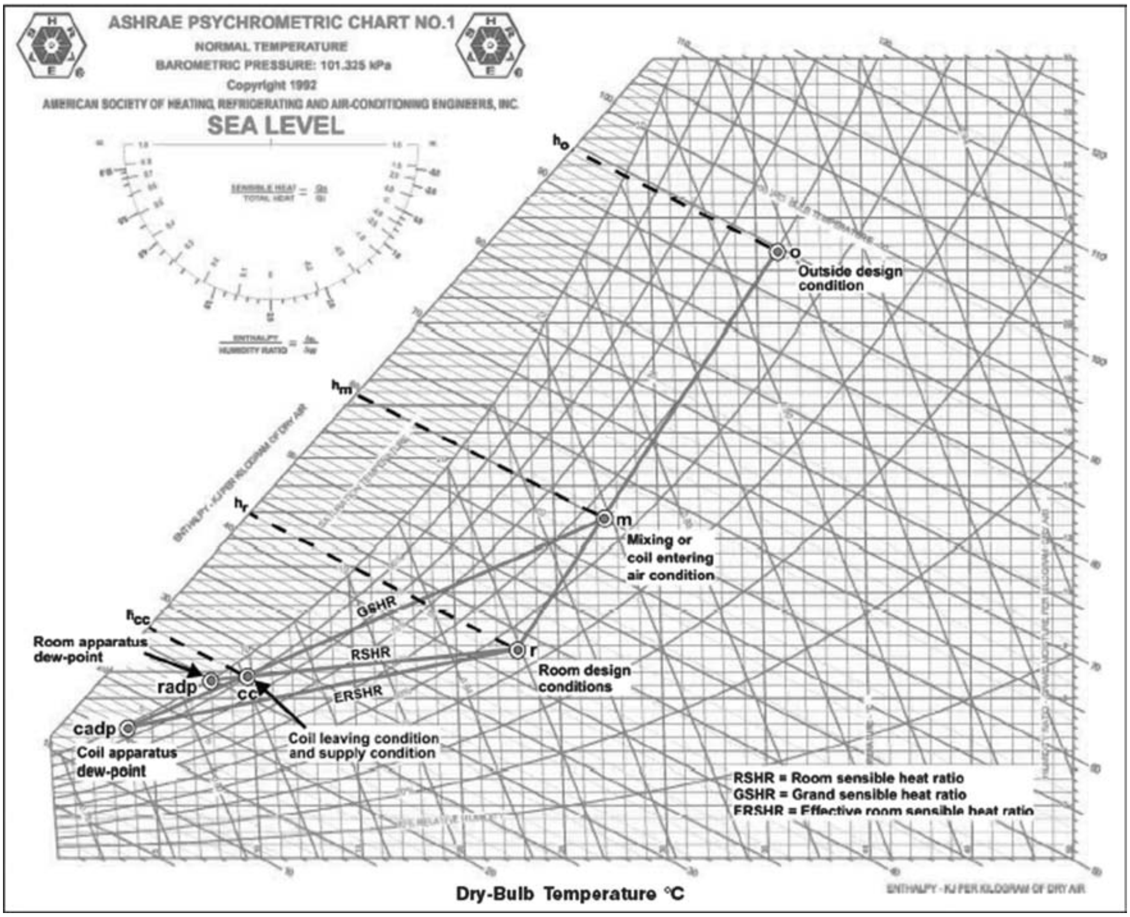


Fig. 3 Psychrometric depiction showing coil bypass factor and sensible heat ratios.

The figure also shows the lines representing the room sensible heat ratio (RSHR), the coil (or grand) sensible heat ratio (GSHR), and the effective room sensible heat ratio (ERSHR). These are defined according to the following equations:

$$\text{RSHR} = \frac{\dot{Q}_{R,S}}{\dot{Q}_{R,S} + \dot{Q}_{R,L}}$$

$$\text{GSHR} = \frac{\dot{Q}_{G,S}}{\dot{Q}_{G,S} + \dot{Q}_{G,L}}$$

$$\text{ERSHR} = \frac{\dot{Q}_{R,S} + B_F\dot{Q}_{O,S}}{[\dot{Q}_{R,S} + B_F\dot{Q}_{O,S}] + [\dot{Q}_{R,L} + \dot{B}_F\dot{Q}_{O,L}]}$$

where the Q_S represent the loads. Subscripts R, O, and G represent room, outside, and grand, respectively, while S and L represent sensible and latent (loads), respectively. The effective sensible heat ratio is interwoven with both the coil apparatus dew point temperature and the coil bypass factor, with the sole intention of simplifying psychrometric calculations. The mass flow rate using the effective RSHR can be computed from the following equation:

$$\dot{m}_s = \frac{(\dot{Q}_{R,S} + B_F\dot{Q}_{O,S})}{\rho C_p(T_r - T_{cadp})(1 - B_F)}$$

where T_r is the indoor (room) design temperature and T_{cadp} is the coil apparatus dew point temperature. The mass flow rate term represents the quantity of air per unit time (kg/s or lb/hr) supplied to the conditioned space. The above equation is not exact because the product of the specific heat (C_p) and the temperature difference was used in lieu of the enthalpy difference. However, the specific heat of air does not change much with temperature in the air conditioning temperature range, and hence the error introduced by the above approximation is negligible.

Heating and Dehumidification

This is also referred to as desiccant (or chemical) dehumidification, which takes place when air is exposed to either solid or liquid desiccant materials. The mechanism of dehumidification in this case is either absorption (when physical or chemical changes occur) or adsorption (when there are no physical or chemical changes). During the sorption process, heat is released. This heat is the sum of the latent heat of condensation of the absorbed water vapor into liquid plus the heat of wetting. The latter quantity refers to either wetting of the surface of the solid desiccant by the water molecules or the heat of solution in the case of liquid desiccant. Dehumidification by solid desiccants is represented on the psychrometric chart by a process of increasing dry-bulb temperature and a decreasing humidity ratio. Dehumidification by liquid desiccants is also represented by a similar line, but when internal cooling is used in the apparatus, the process air line can also go from warm and moist to cool and dry on the chart.

Adiabatic Mixing of Air Streams

This usually refers to either adiabatic mixing of two or more air streams or to bypass mixing. In the former process, two or more streams are mixed together adiabatically forming a uniform mixture "m" in a mixing chamber. In this case, mass, energy, and water vapor balances yield the following equations:

$$\dot{m}_1 + \dot{m}_2 + \cdots + \dot{m}_j = \dot{m}_m$$

$$\dot{m}_1 h_1 + \dot{m}_2 h_2 + \cdots + \dot{m}_j h_j = \dot{m}_m h_m$$

$$\dot{m}_1 W_1 + \dot{m}_2 W_2 + \cdots + \dot{m}_j W_j = \dot{m}_m W_m$$

Bypass mixing, on the other hand, happens usually in air handling units where the air flow is divided into upper hot deck and lower cold deck streams. In a summer air conditioning operation, the upper hot deck acts as a bypass air stream in order to moderate the temperature of the otherwise overcooled air leaving the cooling and dehumidifying coil. In the winter season, on the other hand, the lower cold deck acts as the bypass stream by mixing with the warm air after it has passed over the heating coil (Fig. 4).

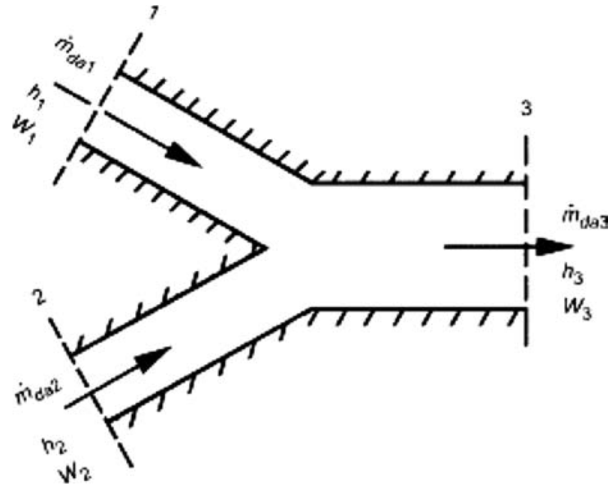


Fig. 4 Adiabatic mixing of air streams.

AIR CONDITIONING CYCLES AND SYSTEMS

An air conditioning cycle is a combination of several air conditioning processes connected together. Different systems are characterized by the type of air conditioning cycle they use. The main function that psychrometric analysis of an air conditioning system achieves is the determination of the volume flow rates of air to be pushed into the ducting system and the sizing of the major system components.

There are generally four extreme climatic conditions that an air conditioning system may face. In summer operation, for example, the dry-bulb temperature of the outdoor air is always high, but the humidity ratio may either be high or low. In hot and humid climates (such as Miami, Florida), the air conditioning system is typically composed of a cooling coil whose surface temperature is below the dew point temperature (see Fig. 5). That way, both cooling and dehumidification can be achieved by the system. In hot and dry climates (such as Phoenix, Arizona), on the other hand, evaporative coolers are typically used.

Winter conditions may have similar extremes, with the dry-bulb and dew point temperatures being both low. In extremely cold conditions (such as Minneapolis, Minnesota), the environment is typically very dry. In this case, the air conditioning system is typically composed of a heating coil and a humidifying device, with the former being located upstream of the latter. The humidifying device may be a spray washer composed of a spray chamber in which a number of spray nozzles and risers are installed. Spray washers are typically used in industrial applications where the device performs the dual function of air humidification

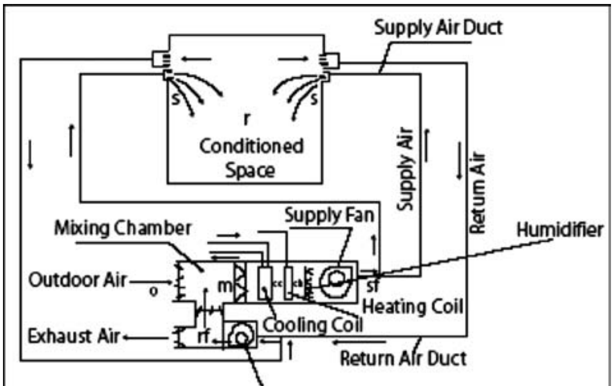


Fig. 5 Basic air-conditioning system.

and cleaning. The washer may have one or more banks of spray nozzles that have the capacity of injecting one to two gpm of atomized water per nozzle into the air stream. Adequate atomization of the water can be achieved by operating at pumping pressures ranging from 20 to 40 psi.[3] Washers typically employ baffles at the inlet air section in order to distribute the air uniformly throughout the chamber. At the exit section, on the other hand, moisture eliminators are employed to prevent carryover water from exiting the chamber into the conditioned space.

In cases where the outdoor air is cool but humid (such as Seattle, Washington), there may not be a need to humidify the air, and the air conditioning system is typically composed of a heating coil only. However, because of the fact that the humidity ratio at low temperatures is low even though the relative humidity may be high, some humidification may be required when the dew point is low during the winter season.

In addition to the four extreme climatic conditions described above, the air conditioning system may also have to operate in the fall and spring seasons. This constitutes a mixed-mode operation that usually requires switching back and forth between heating and cooling based on the value of the outdoor temperature. However, because of the potential energy waste associated with this mode of operation, the system is usually set to operate only if the outdoor temperature goes outside a prespecified wide band in order to minimize the frequency of cycling between the heating and cooling modes.

In order to provide the reader with a flavor of the psychrometric analyses that need to be performed on an air conditioning system, only three of the above systems (hot and humid, hot and dry, and cold and dry) will be described in more detail. The cool-and-humid mode of operation is identical to the cold-and-dry mode except in the absence of the humidifying device.

Summer Hot-and-Humid Mode

Fig. 5 shows a basic air conditioning system capable of both summer and winter mode operations. Fig. 6 shows the psychrometric representation of a hot-and-humid summer mode operation applicable to Miami, Florida. In this mode, outdoor air at State "o" is adiabatically mixed with

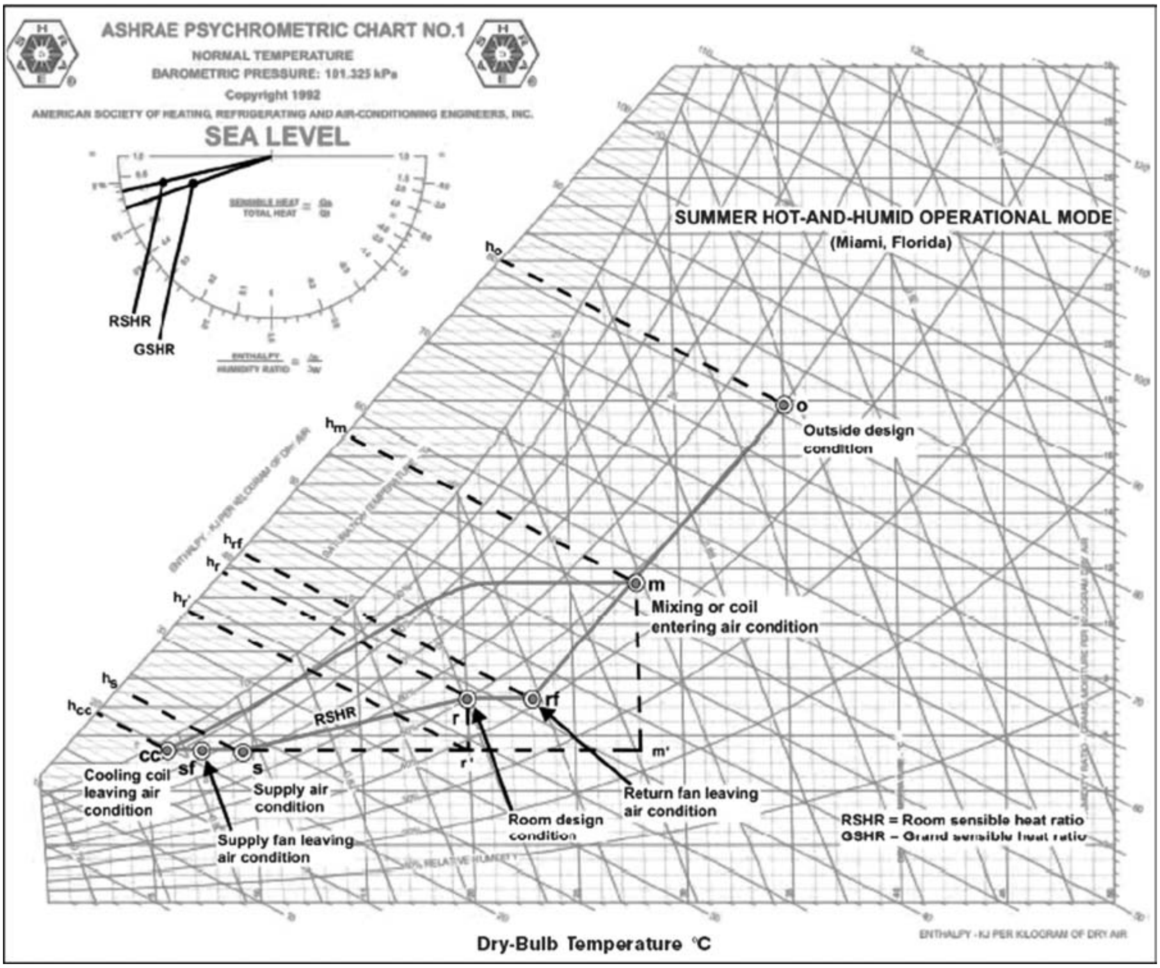


Fig. 6 Psychrometric depiction of hot, humid summer air conditioning system.

recirculated air from the room after passing through the ceiling plenum, the return duct, and the return fan (State "rf") to form the mixed air condition "m," which is also the entering state to the cooling coil. Air is then cooled and dehumidified until it exits the coil at State "cc." After that, air is reheated, due to passing over the supply fan and through the supply ducts, to State "s," the supply condition. The space sensible, latent, and total loads can, respectively, be expressed as follows:

$$\dot{Q}_{R,S} = \dot{m}_s(h_{r'} - h_s)$$

$$\dot{Q}_{R,L} = \dot{m}_s(h_r - h_{r'})$$

$$\dot{Q}_{R,T} = \dot{m}_s(h_r - h_s)$$

where, again, the Q_S represent the load quantities as before and the enthalpies have subscripts corresponding to the respective points identified on the psychrometric chart. Any one of the above equations can be used to compute the supply-mass flow rate. The supply-volume flow rate, on the other hand, is computed with the knowledge of the specific volume at the supply condition v_s. This is expressed as follows:

$$\dot{V}_s = \dot{m}_s v_s$$

The space (room) sensible heat ratio (RSHR) is graphically represented by line "rs" on the psychrometric chart. As for the cooling coil (or grand) loads, the following equations apply:

$$\dot{Q}_{G,S} = \dot{m}_s(h_{m'} - h_{cc})$$

$$\dot{Q}_{G,L} = \dot{m}_s(h_m - h_{m'})$$

$$\dot{Q}_{G,T} = \dot{m}_s(h_m - h_{cc})$$

Again, the subscripts of the enthalpy quantities correspond to state points identified on the chart. The above equations can be used to size the cooling coil. The GSHR is graphically represented by line "m cc." The relative humidity of the air exiting the cooling coil is a function of the coil design, fin spacing, coil surface area, and coil face velocity, among other factors. For a coil with 10 or more fins per inch and four rows of coil, the relative humidity of the leaving air is approximately 93%. For six- and eight-row coils with fin spacing of ten or more per inch, the exiting relative humidities are 96 and 98%, respectively.[10]

Summer Hot-and-Dry Mode

Psychrometric depiction of the processes involved in this mode of operation is shown in Fig. 7. As indicated earlier, the air conditioning equipment in this case is simply composed of a humidifying device where atomized water is introduced into the air stream. Outdoor air at State "o" is typically mixed with return air from the conditioned space after passing through the return ducts and over the return fan (State "rf") to form the mixed state "m." The evaporative cooling process that results from the spray equipment usually follows a constant wet-bulb temperature line, with the air exiting at State "hh." After that, the air passes over the supply fan and through the supply ducts until it enters the conditioned space at State "s." The humidifying effectiveness, η_H, can be expressed either in terms of a temperature deficit ratio or a humidity ratio deficit ratio as follows:

$$\eta_H = \frac{T_m - T_{hh}}{T_m - T_{sat}}$$

$$\eta_H = \frac{W_{hh} - W_m}{W_{sat} - W_m}$$

Special care has to be exercised in analyzing room and grand loads in this mode as the sensible and latent components have opposite signs. For example, the evaporative cooling process employed results in a decrease in the air temperature (sensible cooling), whereas it increases the humidity ratio (latent heating). The process results in a very small change in the air enthalpy (almost zero), as the adiabatic saturation process through the humidifying device follows a constant wet-bulb temperature line.

Winter Mode

There are two possible scenarios in a winter system—one involves the supply of heated air and the other involves the supply of unheated air to the conditioned space. The latter mode is applicable in cases when the outdoor temperature is not too cold that the fan and duct heating may be enough to provide for comfort level temperatures. Because of the similarities between these scenarios, this text will only consider the former scenario for further analysis. As mentioned earlier, for the cold-and-dry mode, there is a need to humidify the air stream. Fig. 8 shows the basic cycle represented on the psychrometric chart for this operational mode. Outside air at State "o" is mixed with recirculated air from the room after passing over the return fan (State "rf"), producing the mixed state "m." Air at this state passes over a heating coil and leaves at State "ch." After the sensible heating process "m ch," air enters the humidifying device where both the dry-bulb temperature and humidity ratio of the air stream are increased, producing state "hh" at the exit of the device. In the humidifying device, either steam or atomized hot water is injected. Air then passes over the supply fan and through the supply ducts to the conditioned space at the supply condition "s."

The amount of heating required by the heating coil typically needs to be coordinated with the heating performed by the humidifying device due to steam or hot water injection. The optimum amount of heating performed by both devices may be hard to pinpoint in an

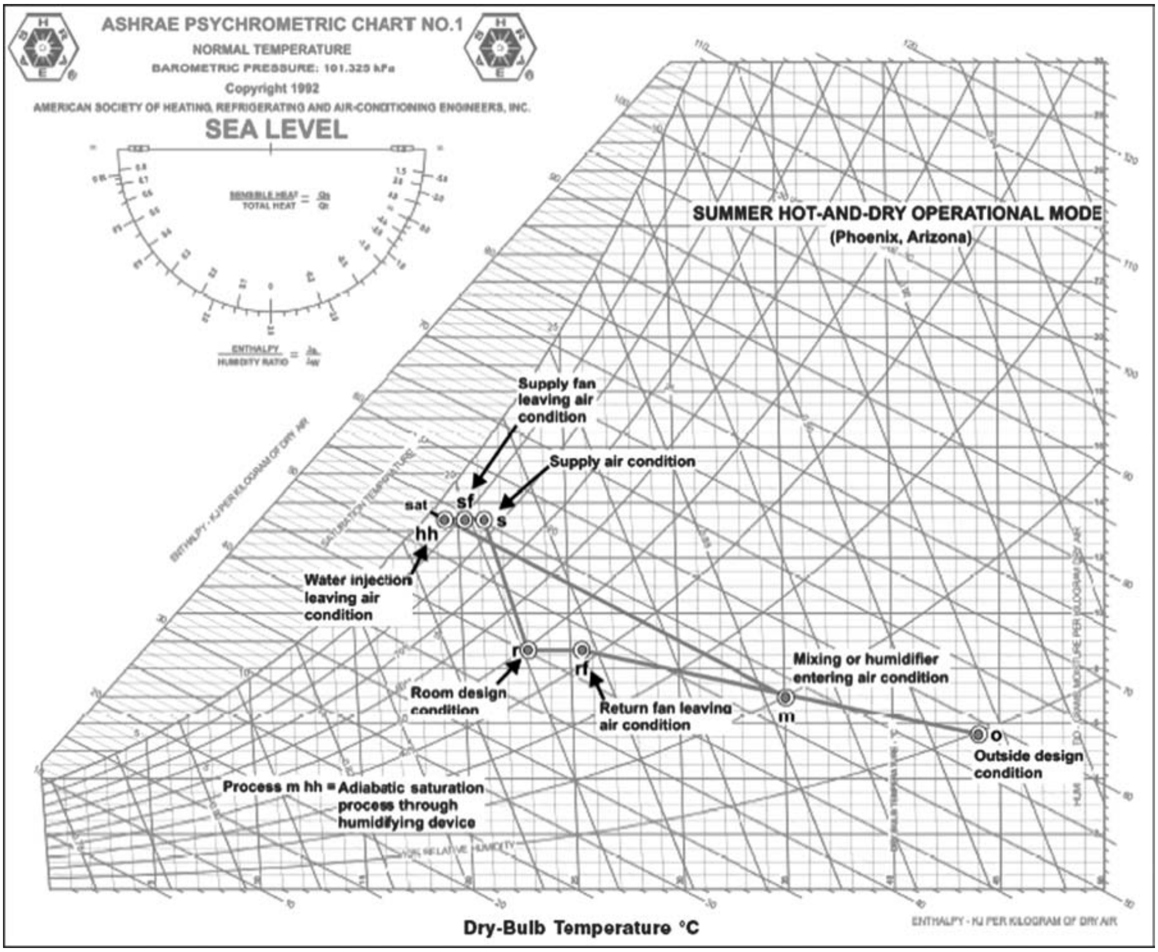


Fig. 7 Psychrometric depiction of a hot, dry summer air conditioning system.

exact way, but a careful choice of the state of air leaving the heating coil may go a long way towards minimizing unnecessary use of heating energy.

The sensible heat ratio for the humidifying device is represented by line "ch hh." The GSHR of the overall system is represented by line "m hh."

In cases when the outside air is humid enough (such as in the Pacific Northwest of the continental United States in winter months), injecting steam or atomized hot water may not be necessary and sensible heating is the only process that may be required.

CONCLUSIONS

This entry presented an overview of psychrometric processes and systems as applied to different operational modes. Proper execution of this stage is crucial in accurately computing the volume flow rates of air through the air conditioning ducts. This phase, in an air conditioning design process, follows the load calculation phase. While the latter phase produces quantities that represent the sensible and latent loads imposed on the conditioned space, the former phase (psychrometric analysis) is capable of incorporating the effects of introducing fresh outside air into the space for ventilation purposes. Cooling or heating equipment sizing has to take into account not only the load imposed on the conditioned space, but also the outside load. By specifying the amount of outside air to be introduced for a specified set of outdoor and indoor conditions, psychrometric analysis enables the HVAC system designer to compute the load imposed on the conditioning equipment (grand load). The analysis is inherently capable of distinguishing between the sensible and latent load quantities of outside and conditioned space (room) air, thus providing a truly insightful picture of how the moisture present should be handled. Psychrometric analysis also enables the designer to account for other smaller loads that may be imposed on the system, such as ducts and fans, in equipment sizing. By identifying the different state points of the air as it passes through the duct system and over the supply and return fans, the volume flow rates of air computed by the analysis become necessarily inclusive of the effects of the ducts and fans in equipment sizing.

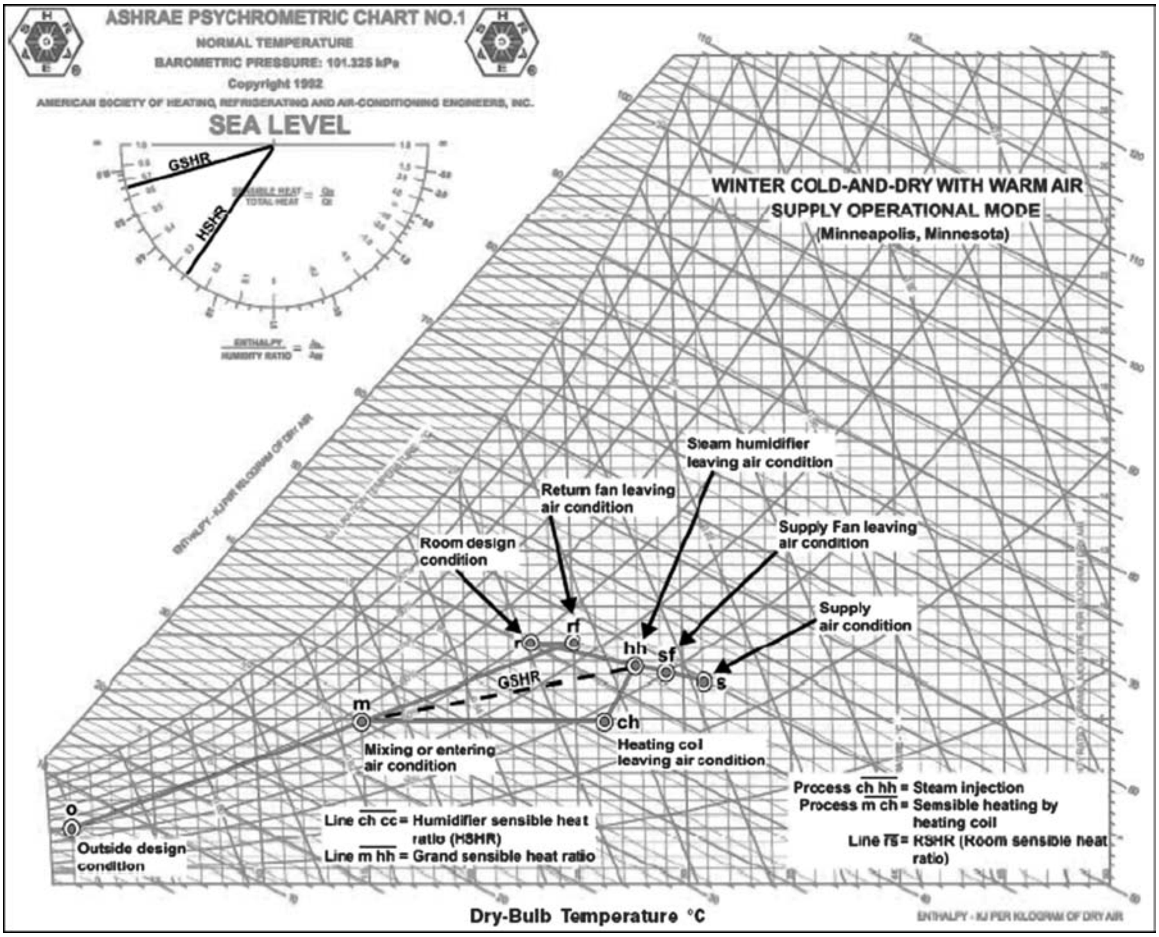


Fig. 8 Psychrometric depiction of a cold, dry winter air conditioning system.

REFERENCES

1. ASHRAE. *Psychrometrics: Theory and Practice*; The American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.: Atlanta, GA, 1996.
2. ASHRAE. *2005 ASHRAE Handbook—Fundamentals*; The American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.: Atlanta, GA, 2005.
3. Clifford, G. *Modern Heating, Ventilating, and Air Conditioning*; Prentice Hall: Englewood Cliffs, NJ, 1990.
4. Hyland, R.W.; Wexler, A. Formulations for the thermodynamic properties of the saturated phases of H_2O from 173.15 to 473.15 K. ASHRAE Trans. **1983**, *89* (Part 2A), 500–519.
5. Hyland, R.W.; Wexler, A. Formulations of the thermodynamic properties of dry air from 173.15 to 473.15 K, and of saturated moist air from 173.15 to 372.15 K, at pressures to 5 MPa. ASHRAE Trans. **1983**, *89* (Part 2A), 520–535.
6. Mago, P.J.; Sherif, S.A. Psychrometric charts and property formulations of supersaturated air. HVAC&R Res. **2005**, *11* (1), 147–163.
7. Nelson, H.F.; Sauer, H.J.; Huang, X. High temperature properties of moist air. ASHRAE Trans. **2001**, *107* (Part 2), 780–791.
8. Sauer, H.J.; Nelson, H.F.; Huang, X. The search for high temperature experimental psychrometric data. ASHRAE Trans. **2001**, *107* (Part 2), 768–779.
9. Stewart, R.B.; Jacobsen, R.T.; Becker, J.H. Formulations for thermodynamic properties of moist air at low pressures as used for construction of new ASHRAE SI unit psychrometric charts. ASHRAE Trans. **1983**, *89* (2A), 536–548.
10. Wang, S.K. *Handbook of Air Conditioning and Refrigeration*; McGraw-Hill Book Company: New York, 1993.

Public Policy for Improving Energy Sector Performance

Sanford V. Berg
Director of Water Studies, Public Utility Research Center, University of Florida, Gainesville, Florida, U.S.A.

Abstract
Public policy evolves in response to a number of forces, including public perceptions and special interest campaigns. Policy influences the basic conditions in the energy industry, in market structures, in corporate behavior, and in energy sector performance. Government intervention is often justified on the basis of market imperfections (market power and information problems), market failures (pollution), and income distributional concerns. It affects the performance of the energy sector through regulation, antitrust, environmental mandates, and other mechanisms.

INTRODUCTION

Public policy establishes the legal constraints facing decision-makers and determines the jurisdictional responsibilities of different levels of government. The basic rationale is that market imperfections (market power and information gaps) and market failures (such as pollution) require some form of government intervention. Energy policies tend to address three broad areas: market structure, corporate behavior, and sector performance. While antitrust regulation addresses mergers and anticompetitive behavior in the economy, energy sector economic regulation tends to address those elements of the market structure (including the supply chain) that are viewed as natural monopolies. Government intervenes when economic or social problems catch the attention of those involved in political processes. Politicians then craft legislation and create administrative agencies to implement the laws. The courts rule on the legality of different arrangements when laws come into conflict or are challenged in terms of constitutionality. In addition to antitrust agencies and energy sector regulators, environmental agencies have also been created to address the ecological and health impacts of energy production and distribution. Thus, pressure for changes in public policy builds as energy sector performance falls short of expectations. The relative roles of markets and government shift when citizens lack confidence in current institutional arrangements and when new social concerns arise.

Keywords: Market power; Environmental impacts; Regulation; Incentives; Restructuring.

CREATION OF LAWS AND RULES

The standard justification for government intervention in the economy is to "improve performance." Of course, different dimensions of performance will tend to be given different weights, according to how citizens view current social and economic outcomes. Energy policy is particularly sensitive to public opinion because the sector significantly affects citizens; in addition, energy suppliers are often politically powerful. The range of concerns means that political coalitions (based on regional alliances or ideological predispositions) form around issues and support policy initiatives that meet their concerns. Some groups are concerned with social justice (or fairness), particularly regarding the effect of energy prices on low-income citizens. Others worry about environmental impacts associated with energy and seek investments in research and development (R&D) and conservation to reduce those impacts. Current suppliers tend to focus on financial success and seek programs that reduce risks and increase returns. Potential entrants seek a level playing field so that incumbents do not benefit from artificial advantages. Customers understand that if firms are not efficient, prices will be higher than they need to be. In addition, particular customer groups are interested in whether they are subsidizing other groups, so the price structures also matter.[1]

All of these interest groups communicate their concerns to politicians. In addition, the media (newspapers, television, and, increasingly, the Internet) shape citizen attitudes by providing information about the implications of various policy options. Policy options include changes in sector taxes, subsidies, entry conditions, environmental mandates, and other regulatory rules. Because managers of private and publicly owned enterprises view government as constraining decision-making, they will track potential policy changes and lobby against actions that seem to harm the interests of those enterprises. Another term for self-interested policy promotion to obtain favorable treatment from governmental authorities is "political rent-seeking."[2]

The policy-making process involves convincing key stakeholders that the change is (or is not) needed. During the public debate, different groups provide information to policy-makers on alternative approaches to resolving issues. Public policies often concentrate benefits while dispersing costs over a number of groups. For example, large corn farms benefit from subsidies for ethanol. For groups with high per capita benefits, political lobbying will be intense. This pattern means that public intervention sometimes results in greater aggregate costs than social benefits, as when special interests gain protection from developments that threaten their positions of privilege. In other instances, new laws create win–win outcomes that can enhance energy sector performance.

Although public policy is partly based on legislation, it is also developed by government ministries, departments, and agencies established by the law and responsible for creating rules based on the policy. For example, energy sector regulatory commissions have the task of implementing the law by developing procedures that meet the objectives established by political authorities. Each nation has a unique history and legal structure, so public policies have tended to differ across countries.[3] Under federal systems such as in the United States, India, Brazil, and Russia, state and national authorities generally have a division of labor that gives states jurisdiction over local energy activity (such as electricity distribution). However, the division of responsibility in the area of energy depends on constitutional authority and on the interaction between public expectations and sector performance over time. Jurisdictional disputes are likely to arise when the status quo is challenged by one group or another.[4]

RATIONALE FOR INTERVENTION

Competitive markets often promote efficient outcomes, but problems occasionally arise when the market does not fully account for all benefits and costs or when it produces an unreasonably inequitable outcome. Government intervention is often justified on the basis of these market imperfections, market failures, and income distributional concerns.[5]

Market imperfections include market power and information asymmetries. Market power arises if there is a single supplier or if suppliers collude to earn financial returns far greater than required to attract investors, and consumers face high prices. Two major policies address market power: sector regulators deal with natural monopoly and antitrust authorities attempt to limit the artificial attainment (and exercise) of market power. Of course, the scale economies associated with natural monopoly can change. When technological change alters the production economies on which previous market structures had been predicated, public policy may change in response to competitive possibilities. The transition to new market structures may require that regulators develop rules for access to essential facilities (as when generators need access to the transmission grid).

Another market imperfection involves information problems. For example, do consumers have full information about the life-cycle costs of energy-using appliances? What time horizons do home buyers use when considering appropriate levels of insulation? Product labeling requirements and energy efficiency mandates are based on an assumed lack of consumer information.

Market failures arise when there are environmental impacts. In such cases, those who produce and consume the product are not the only ones affected by the transaction. The energy production chain (exploration, extraction, transport, refining/generation, transmission, distribution, and retailing) involves a number of sensitive areas. Thus, air pollution, water pollution, and land use impacts (plant siting and the placement of high voltage lines) have come under the purview of energy regulation. (See the entry on *Environmental Regulation*.) Environmental issues can also be examined in the context of common property resources—who "owns" the clean air? Another potential market failure arises in the context of fuel diversity: do competitive generation markets yield the optimal portfolio of fuel types (including renewable resources) so end users do not bear the excessive risk of price increases under particular fuel price scenarios?

Income distributional concerns, such as universal access to electricity, involve low-income or high-cost (rural) consumers. Nations generally devise subsidies of one form or another to promote network expansion into high-cost geographic areas. In addition, costs may be covered by public funds or cross-subsidies to promote usage by those otherwise unable to afford what is generally viewed as a necessity. Of course, not all subsidies are well-targeted, and once a group is receiving subsidies, governments find it difficult to cut them off even when circumstances change.[6] The ethical objective of social justice is difficult to define and to quantify.

All three rationales for intervention relate to sector performance that does not match citizen expectations. Besides performance, citizens also care about the process used to develop and implement policy. Was the process transparent and did it involve citizen input (participation)? Procedural fairness affects citizen perceptions regarding the legitimacy of public policy. The ultimate form of intervention (and potential inefficiencies) becomes a matter of the relative political power of different stakeholders. However, the key point is that public policy is not shaped in a vacuum. There is a rationale for intervention even when the specifics of the adopted policy include elements that favor particular groups or create inefficiencies.[7]

FACTORS AFFECTING POLICIES

The figure is a flow diagram that shows how public policy responds to sector performance.[8] The behavior of firms (in terms of pricing, cost-cutting, provision of services, and network expansion) follows from the market structure and rules affecting behavior. Ultimately, public perceptions are influenced by sector performance, including the adoption of innovations, network reliability, environmental impacts, and other outcomes. Indicators of sector performance include returns that are commensurate with risks, production efficiency, introduction of new technologies, safety in the production process, and sound environmental stewardship. Thus, public policy changes arise from weak sector performance relative to what is deemed possible. Perceptions are affected by information presented in the media and by opinion leaders (Fig. 1).

The diagram shows how social objectives and priorities influence the creation and operation of regulatory commissions, antitrust agencies, product information regimes, and environmental agencies. The law generally specifies the policy objectives and the approaches to be utilized by the agencies implementing public policy. However, regulatory commissions and other government agencies have a significant impact on the ultimate rules that are adopted, with court challenges serving as another influence on policy implementation. The boxes in the figure represent fourteen factors that are part of the policy environment, which can be understood as follows:

1. Objectives and Priorities: The broad economic and social objectives of citizens include freedom, equality, justice, high living standards, and technological advancement.[5] In the context of energy, political leaders attempt to discern (and shape) what citizens want from infrastructure sectors. Social values may reflect a consensus or be deeply divisive and lead to dramatic shifts in public policy. Events such as an oil crisis, a serious accident (like Three Mile Island), or a natural disaster can also trigger changes in public priorities and a willingness to move from the status quo.
2. Institutional Conditions: These represent the starting point for policy development. Mediated by the press and political leaders, they are the context in which agencies are created and "reformed."[9] Institutional conditions can be characterized by the extent of consensus between the legislative and executive branches of government, the judicial capabilities (and consistency in legal decisions), agency administrative capacity (whether highly professional or heavily politicized), informal norms, and formal rules. Informal norms are the customs that mold day-to-day behavior. When corruption is accepted behavior, firms and customers are subject to arbitrary treatment, which increases risk and raises costs. Formal rules are embedded in the legal framework in which organizations operate.
3. International Experience: History provides evidence regarding which institutions and policies have failed and which have been successful. Experience from other countries (or states within a nation) represent "natural experiments" regarding impacts of alternative energy policies and provide lessons for policy-makers. Thus, infrastructure performance across jurisdictions and over time yields a yardstick for determining whether performance is subnormal. Dubash[6] outlines how different national policies have attempted to address a broad range of social and economic objectives.

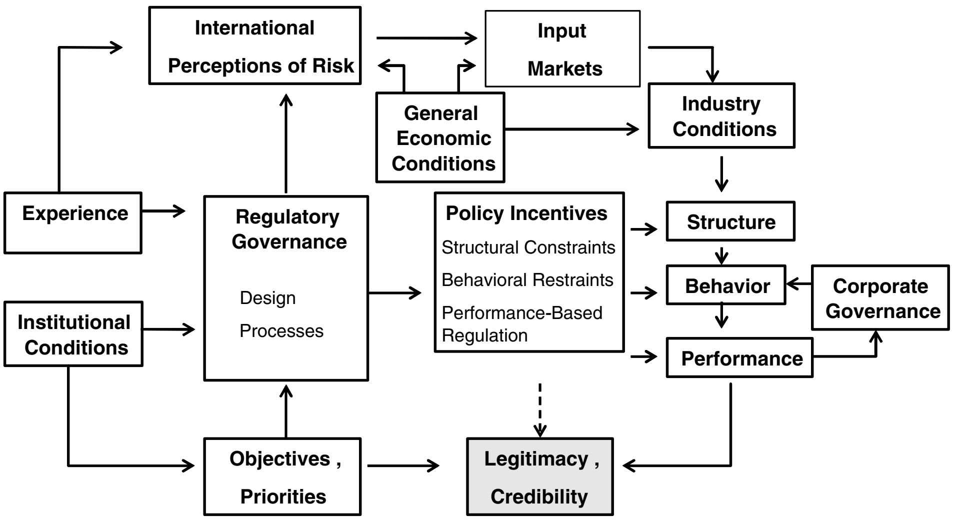


Fig. 1 Determinants of public policy and sector performance.

4. International Risk Perceptions: International perceptions regarding national security, energy independence, trends in the exchange rate, and concerns over cross-national environmental impacts affect investor attitudes and thus the cost of capital (reflected in required interest rates on bonds and expected returns on equity investments) but tend to be beyond the control of sector regulators or antitrust authorities. An exception concerns the predictability of government rules—an area where public policy does affect international perceptions. If the rules of the game are continually changing, investors are less willing to commit substantial funds to projects whose cash flows depend on a stable regulatory environment.[10]
5. General Economic Conditions: Macroeconomic factors are also outside the purview of those enforcing rules in the sector. Trends in the exchange rate will be affected by monetary and fiscal policy. Clearly, income growth has important consequences for energy demand and for the profitability of investment in the sector.
6. Input Markets: The presence or absence of key inputs determines earnings from investment. For example, operational effectiveness depends on the skills and incentives of managers and engineers. In addition, domestic and international capital markets affect the supply and demand of capital, thus the cost of capital. Similarly, the training of skilled engineers and thoughtful managers affects the availability of these key inputs. The availability of natural resources is another aspect of input markets affecting performance. Dependence on imported high-priced energy affects the costs that are ultimately borne by consumers. After considering demand growth and production technologies, these inputs create the basic conditions in the energy sector.
7. Basic Industry Conditions: Basic conditions include production technologies, input prices (fuel, capacity, materials, and labor), demand patterns (and growth), and ownership patterns. When the scale economies for generation are large relative to the market size, a natural monopoly may exist because having a single supplier can be least-cost. Such a situation leads to a market structure with a single (often vertically integrated) electric utility. Thus, a regulatory commission might be created to constrain the exercise of market power by the single supplier. If demand grows substantially or small-scale production becomes economical (as with combined cycle gas turbines or distributed generation), generation can become potentially competitive and lead to pressures by potential suppliers to lower entry barriers.[11]

   Common property resources (clean air or water) create situations that invite government intervention. To some extent, technical standards (such as those developed under the auspices of the Institute of Electrical and Electronics Engineers) or energy efficiency standards mandated by law represent part of the information structure constraining utility decisions. Basic industry conditions change with national expenditures on energy R&D as new production technologies are developed and introduced. R&D might be performed in industry collaborative ventures (such as the Electric Power Research Institute in the United States), universities, and product suppliers. R&D funding comes from a variety of public and private sources. For nations with domestic natural gas production, regulation of gas transmission and distribution pipelines has often been added to the responsibilities of the entity having electricity oversight. Private and public ownership is another factor affecting downstream activities. Changes in basic conditions affect sector regulation in other ways. For example, the development of nuclear power precipitated the creation of specialized agencies responsible for setting standards for safety and public health aspects of the technology (including storage and disposal). National agencies like the Atomic Energy Commission (now the Nuclear Regulatory Commission) in the United States illustrate this trend. In addition, international agencies (International Atomic Energy Agency) have been created to address transnational issues and to monitor developments.
8. Regulatory Governance: Stepping back from how basic conditions affect the structure, behavior, and performance of an industry, we need to recognize the important role of the features characterizing the agency responsible for implementing public policy. Regulatory governance includes the legal mandate given to a government agency, the resources available for policy implementation, the organizational design of the agency, and the processes adopted by the agency, all of which affect regulatory activities. In addition, if there is no clarity in terms of which agency is responsible for implementing particular policies, public policy is likely to be compromised by intragovernmental rivalries. Resulting policy incentives affect the structure of the energy sector, the behavior of energy suppliers, and the performance of the industry. Baldwin and Cave[2] survey the key features of regulatory systems. Incentives can include structural constraints,

behavioral restraints, and performance-based initiatives (discussed below).

9. Policy Incentives: These can include taxes and subsidies that discourage and encourage a variety of activities. Regulatory incentives include constraints on market structure, restraints on corporate behavior, and performance-based rewards and penalties. The significant policy issues involve designing incentives that promote cost-cutting, service quality, and network expansion. In some cases, changes in public policy can significantly lower the cash flows that can be obtained from productive assets. For example, if allowing additional entry into the production of electricity means that "old" plants are operated for fewer hours per year, the net cash flows associated with those plants decline. Analysts can debate whether (and when) regulatory policy changes could have been anticipated and factored into investment decisions. On the practical side, if a restructuring initiative is adopted (to unbundle what was traditionally a vertically integrated industry), policy-makers generally try to address the issue of how to deal with the lost economic values stemming from the policy change. For example, some U.S. states have imposed competitive transition charges so consumers bear some of the burden of moving to a new market structure and a new regulatory framework (sometimes labeled "stranded costs").
10. Market Structure: For a homogenous product like electricity, market structure can be characterized in terms of entry conditions, the degree of vertical integration, and other factors. Government policies greatly influence the number and size distribution of suppliers through merger policy and the creation of franchise territories. Up until the last few decades, the electricity sector generally involved local monopolies, where entry was restricted by law. In the United States, these firms were often privately owned but regulated; in many nations, the governance system for publicly owned utilities was not very transparent, with policy-making, policy implementation, and operations often conducted within a government ministry. Policy-makers in many nations are currently exploring the extent to which competition can substitute for regulation. Of course, effective competition requires a sufficient number of firms that are operating independently from one another so that they become "price takers" (and lack market power). There is substantial evidence that competition can substitute for regulation when economies of scale and multiproduct economies make entry feasible. Thus, policies affecting market structure, vertical integration, and corporate ownership shape the evolution of the energy sector.
11. Corporate Behavior: Public policy also creates incentives involving behavioral restraints. These incentives are related to price, quality-of-service requirements, and mandates for system expansion. Antitrust laws attempt to address price-fixing, undue price discrimination, market foreclosure, and other issues. Sector regulators use cost of service (or rate-of-return regulation), price and revenue caps, and other mechanisms for constraining prices. For example, electricity regulators often include some fuel cost adjustment in the price mechanism to pass unavoidable cost increases (and decreases) through to consumers. The customer's bill is changed when the actual cost of fuel at the supplier's generating stations varies from a previously specified unit cost. Of course, such clauses insulate suppliers from input price fluctuations, reducing their incentive to seek low prices or to reduce the risk through hedging instruments. Other automatic adjustment clauses fund conservation programs or environmental programs (including "green" energy). Every regulatory rule has some impact on corporate behavior.
12. Sector Performance: Another set of policies affect sector performance. In the case of energy, regulators may mandate reliability requirements, network expansion targets, or profit-sharing plans. In the case of sharing plans, instead of establishing strict rate-of-return limitations (with associated weak incentives for cost containment), regulators may split the savings from performance improvements and allow investors to benefit from above-normal productivity gains. Meeting targets is often encouraged through performance-based ratemaking (PBR), which fits into a broad category of rate-setting mechanisms that link rewards to desired results or targets by setting rates (or rate components) for a given time according to external indices rather than a utility's actual cost of service. This type of regulation gives utilities better cost-reduction incentives than cost-of-service regulation. In developing performance standards for a PBR plan, a regulatory agency attempts to understand the utility's historic performance in order to develop an appropriate baseline for yardstick comparisons. The regulator collects information on service quality, determines the areas where cost savings may be realized and quality may be improved, and then develops measures for benchmarking performance. Brennan, Palmer,

and Martinez[4] survey the evolution of performance in the electricity industry and the issues confronting policy-makers.

13. Corporate Governance: Traditionally, activities internal to a firm are not micromanaged by regulators, but public perceptions about potential problems have changed in recent years. Corporate governance involves the allocation of decision rights within the firm (hierarchical vs team decision-making), the design of pay plans that compensate people for high levels of effort and performance, the development of performance evaluation mechanisms that monitor the effectiveness of internal reward systems, and dissemination of material about corporate activities. Policy may require that suppliers introduce reporting procedures that limit how insiders might adversely affect investor interests and sector performance. This factor has led to substantial revisions in reporting requirements in a number of nations (including the United States).
14. Legitimacy and Credibility: Ultimately, citizen acceptance of public policies, industrial outcomes, and the political processes that lead to those policies depends on how well industry performance matches shared national objectives. If there is no social consensus (or shared political vision), the legitimacy of the system is continually called into question. The lack of consensus could stem from disagreement regarding facts (whether global warming is occurring and whether industrial development causes it) or lack of consensus regarding values (whether creation of industrial jobs for low-income citizens is more important than devoting resources to improving environmental amenities). In either case, controversy is endemic in energy policy, where different stakeholders seek to influence policy outcomes. Another element affected by sector performance is the credibility of the system to private and public investors. Private investment is a voluntary activity and investors will shy away from projects, companies, or nations whose returns are low relative to the risks. Public investment (through legislative budget outlays or through cash flows to municipal utilities) is also affected by utility performance. If there is evidence of corruption or weak performance, pressures for change build up.

Thus, the figure depicts the circular dynamics of the larger decision environment in which government policy-makers, regulators, energy firms, and public and private sector investors operate and interact. Causation is seldom unidirectional, but changes in basic conditions generally result in altered market structures, corporate behavior, and sector performance. Similarly, changes in public perceptions can lead to significant public policy initiatives.

INFORMATION AND POLICY IMPLEMENTATION

Public policy can be designed to slow down the effects of changes in basic industry conditions or to speed up the system's response. Regulators have only limited information about a firm's commercial activities and its opportunities for cost containment. Regulatory institutions and incentives are partly designed in recognition of this information problem. Also, as noted, technological change and new national priorities can require modifications in regulatory rules and incentives. A nation's institutional endowment (including judicial, legislative, and executive institutions) partly determines the appropriate balance between rigid regulatory rules and flexibility to respond to changing basic conditions.

Regulatory Processes

The creation of a (relatively) independent regulatory commission is one mechanism for insulating those implementing public policy from daily political pressures, and in doing so, balancing the interests of customers, investors, suppliers, and current political authorities. In the context of regulation, a formal hearing procedure provides an avenue for stakeholder participation. A judicial or adjudicatory role is certainly appropriate when the agency needs to determine compliance with specific rules or evaluate behavior. Enforcement proceedings and complaint cases are examples for which this process is appropriate. However, hearings are basically "trial type" adjudicatory processes with a major defect—they tend to emphasize process instead of substance. In such a fact-specific setting, the stability and certainty of the process can cause all parties to overlook the substance of the outcome.

Some jurisdictions have emphasized informal processes for resolving issues. For example, as the issues become more complex, alternative dispute resolution procedures can be effective in identifying win–win options. An "all-party settlement" process represents an alternative way to gain stakeholder buy-ins. In this setting, a negotiated outcome reached by stakeholders is adopted if all parties support the outcome, the public interest is adequately represented, no part of the agreement abrogates any other rule or prior decision, and a factual record is developed to enable implementation.

Reversals or inconsistencies can undermine a regulator's credibility and hence effectiveness. Legislative policy intentions are usually general enough that

Table 1 Decisions and outcomes subject to regulation

Decision	Regulatory oversight
Profit level	"Excessive" returns addressed
Price structure	Rate design and prices to customer groups considered
Product mix	Accounting separations if entry into complementary services
Production process	Environmental mandates and R&D incentives established
Product quality	Reliability standards promulgated
Place	Facilities siting subject to review
Purchasing	Cost containment monitored
Promotion	Advertising and consumer information examined
Planning	Forecasts and plans placed into public record
Partnerships	Mergers and strategic alliances reviewed
Portfolio	Mix of debt and equity (corporate leverage) monitored
Public offerings	Public ownership vs private participation

regulators will need to balance competing goals and values in specific situations. As long as the public record provides clear evidence regarding the facts associated with the issues at hand, the judgment calls of regulators will at least be based on information that others can examine and verify.[7] Without such openness in the process, stakeholders supporting alternative policies will have a case for questioning regulatory decisions. Effective agencies establish procedures that promote accountability, credibility, and legitimacy. Transparency is one mechanism for addressing all three of these aspects of the regulatory process.

Regulatory Activities and Impacts

Effective regulation of prices involves a variety of factors, including the following:

1. Licensing specifies operating standards that affect costs/tariffs.
2. Performance standards for quality and reliability have cost and tariff implications.
3. Monitoring data on costs, revenues, and performance is essential for tariff determination.
4. Tariff setting determines revenue sufficiency for operating costs, capital costs (returns, asset values, deferred loans), etc.
5. Uniform accounting systems provide comparable cost data (generation, transmission, distribution) for tariffs.
6. Arbitration among firms and consumers is needed to settle disputes about tariffs access.
7. Management audits promote cost and tariff reduction via yardstick comparisons.
8. Human resource policies have operating costs and tariff implications.
9. Reports on costs and tariffs emphasize current and future performance and efficiency, both for individual firms and the sector as a whole.

These factors reflect how the implementation of public policy affects corporate decisions. The table lists some of the corporate decisions that might be constrained by public policy. Each of the listed items can be subject to regulatory rules or to statutory limitations and underscore the wide range of issues that arise in the context of energy industries (Table 1).

CONCLUSION

Public policy evolves in response to a number of forces, including public perceptions and special interest campaigns. Policy influences the basic conditions in the energy industry, market structures, corporate behavior, and energy sector performance. Additional factors that influence public policy and the types of agencies developed to implement policy include institutional features of the nation, lessons learned from others, economic growth, and the priorities placed on different performance outcomes. To the extent that actual industry performance falls short of what citizens view as possible or desirable, pressures build for changes in public policy (often labeled "policy reform"). Typical policy objectives include (but are not limited to) the following:

- Constraining the exercise of monopoly power by incumbent suppliers
- Stopping subsidies or creating tax incentives for the sector and discouraging or encouraging particular activities (which affects government deficits)
- Optimizing the structure of the sector (encouraging competition where feasible)

- Providing incentives for operating efficiency and improving reliability
- Promoting least-cost system expansion (incentives for new investment)
- Stimulating innovation and energy conservation

Even when subsidy programs are continued (like those benefiting particular fuels), supporters describe them in terms of how the programs might contribute to meeting other broad energy objectives. Achieving desired outcomes requires some prioritization because policy-makers must make political and economic trade-offs on a continual basis. Here, the focus has been on electricity and the relative roles of markets and government, but similar issues arise in other segments of the energy industry.

REFERENCES

1. Newbery, D.M. *Privatization, Restructuring, and Regulation of Network Industries*; MIT Press: Cambridge, MA, 1999.
2. Baldwin, R.; Cave, M. *Understanding Regulation: Theory, Strategy, and Practice*; Oxford University Press: Oxford, 1999.
3. Armstrong, M.; Cowan, S.; Vickers, J. *Regulatory Reform: Economic Analysis and British Experience*; MIT Press: Cambridge, MA, 1999.
4. Brennan, T.; Palmer, K.L.; Martinez, S.A. *Alternating Currents: Electricity Markets and Public Policy*; Resources for the Future: Washington DC, 2002; xi-210.
5. Greer, D.E. *Business, Government, and Society*; Macmillan: New York, 1993; xviii-606.
6. Dubash, N.K. *Power Politics: Equity and Environment in Electricity Reform*; World Resources Institute: Washington, DC, 2002; xv-175.
7. Viscusi, W.K.; Vernon, J.M.; Harrington, J.E. Jr *Economics of Regulation and Antitrust*; MIT Press: Cambridge, MA, 2000; xxvi-864.
8. Berg, S.V. Sustainable regulatory systems: laws, resources, and values. Utilities Policy **2000**, *9*, 159–170.
9. Levy, B.; Spiller, P.T. The institutional foundations of regulatory commitment: a comparative analysis of telecommunications regulation. J. Law, Econ. Organ. **1994**, *10* (2), 201–246.
10. Kessides, I.N. *Reforming Infrastructure: Privatization, Regulation, and Competition*; World Bank and Oxford University Press: Oxford, 2004; xv-306.
11. Hunt, S. *Making Competition Work in Electricity*; Wiley: New York, 2002.

Public Utility Regulatory Policies Act (PURPA) of 1978

Richard F. Hirsh
Department of History, Virginia Tech, Blacksburg, Virginia, U.S.A.

Abstract
The Public Utility Regulatory Policies Act (PURPA) of 1978 contributed to the restructuring the American electric utility system. It did so by providing incentives for nonutility companies to produce power using energy-efficient generating equipment and renewable-energy resources. By making electricity as cheaply as existing utilities in many cases, the new power companies injected competition into a formerly said, monopolistic industry. The law's unintentional consequences motivated policy makers to consider the value of introducing other free-market principles, some of which became elements of the Energy Policy Act of 1992.

INTRODUCTION

As one part of President Jimmy Carter's comprehensive energy strategy, the Public Utility Regulatory Policies Act (PURPA) of 1978 sought to encourage the efficient use of electricity and stimulate the production of power in nontraditional ways. For the first time, PURPA gave special privileges to nonutility companies that produced electricity using fossil fuels less wastefully than utilities and to firms that employed renewable resources. In the process, the law removed barriers that had previously hindered entry into the generation sector of the power business.

More important, some of the new nonutility producers sold electricity at prices that compared favorably to those offered by regulated power companies. By doing so, they challenged the status of utility companies as natural monopolies—companies that deserved special privileges because they could deliver services more cheaply and efficiently only if they remained in a noncompetitive business environment. The law's implementation further helped spur the introduction of free-market principles, and it forced some policy makers to reconsider the merits of traditional regulation. It, therefore, played a major role in efforts that led to deregulation and restructuring of the American electric utility system.

ORIGIN OF LAW DURING THE ENERGY CRISIS

Even though the so-called energy crisis had occurred more than 2 years earlier, newly inaugurated President Jimmy Carter felt much needed to be done to fix America's energy infrastructure.[1] He recalled how Arab members of the Organization of Petroleum Exporting Countries had imposed an embargo on the shipment of oil to the United States and other countries that supported Israel in its October 1973 war. After the embargo, oil flowed again, but at greatly increased prices that wreaked havoc on the economy and everyday life. Automobile drivers, who had grown accustomed to cheap and easily available gasoline found themselves on long lines for expensive and sometimes rationed fuel. Fortunately for these consumers, the lines disappeared fairly quickly, though gasoline did not return to preembargo prices. The next 2 years saw relatively stable prices for all energy supplies, and many Americans became complacent about energy policy. But President Carter worried about the increasing amount of imported petroleum being used by the United States—rising from 36% in 1973 to 42% in 1976. As his first major initiative after taking office in January 1977, Carter promoted an energy policy that would make the country more energy efficient and less dependent on foreign fuel.

Hoping to rally Americans to fight a "moral equivalent of war," Carter sought to win support for his National Energy Plan. The plan offered tax credits to people who employed energy-efficiency measures in homes and businesses, realizing that conservation was "the quickest, cheapest, most practical source of energy."[2] It also included a set of taxes on those who wasted energy, such as owners of gas-guzzling cars that failed to meet fuel-economy standards. Another proposed tax would have added 5 cents to the price of a gallon of gasoline each year that total gasoline consumption exceeded government-established goals. Additionally, Carter sought to remove price controls on oil and natural gas, which would have allowed the fuels to become more expensive and which would have encouraged users to be more efficient due to economic self interest. He also proposed new measures to spur nuclear power and the use of domestic coal rather than imported oil.

Keywords: Energy legislation; Deregulation; Restructuring; Wind turbines; Photovoltaics; Renewable energy; Cogeneration; Qualifying facilities.

Though legislators may have lauded the president's goals, they proved unwilling to accept all the president's measures, especially those that would have called for potential sacrifice and hardship for their constituents. Instead, they took more than a year to ponder the president's plan and passed a watered-down version of it in the form of five specific laws. At a signing ceremony in November 1978, Carter admitted that the laws did not accomplish everything he had hoped. Nevertheless, they constituted a first step toward a more sound energy policy.

FOCUS OF LAW

Of the five weakened measures, the PURPA had already been viewed as the least consequential. Consisting of 61 pages of text and 78 sections, the law focused on encouraging greater efficiency in the use of electricity.[3] It did so by calling attention to the way electric utility companies charged customers for electricity. Consequently, industry insiders earlier referred to it as the "rate reform" bill. Typically, customers (especially residential customers) paid higher prices per unit of electricity (a kilowatt-hour, or kWh) for the first increment of energy they used each month. This greater cost generally included a large amount of the fixed customer and capital costs associated with the production and distribution of electricity. Subsequent increments cost less per kilowatt-hour, representing only the cost of energy needed to generate the power. This "declining block" rate structure had been used for decades and was popular because it often encouraged customers to use more power, since later increments became relatively cheap. During an age of declining costs for the power industry—an age that lasted from the beginning of the twentieth century until around 1970—the rate structure made perfect sense. But with energy prices increasing, PURPA required utilities and state regulators to reconsider (but not necessarily replace) these rate structures and design new ones that encouraged wise use of electricity. Lobbyists for the utility industry viewed the law as tame, since it did not mandate too much change from the status quo.

SECTION 210

With PURPA's focus on rate reform and with concerns over more severe provisions of other laws that came out of the National Energy Plan (such as the requirement to shift from oil and natural gas to coal for producing power), utility executives and lobbyists paid little attention to the apparently inconsequential Section 210 of PURPA. Dealing with production of power from small-scale generators, the section reflected President Carter's hopes to produce more electricity through unconventional means, thus reducing the amount of oil, coal, or natural gas needed by traditional utilities to make electricity. As one promising approach, a cogeneration facility looked like a tiny utility power plant, but with a twist. It heated water using fossil fuels, and the steam went through a turbine that turned a generator, which then produced power. Instead of venting the waste heat into the environment, like utilities did, a cogeneration plant employed the heat for industrial processes (such as for manufacturing paper or chemicals) or for space conditioning. It, therefore, achieved a higher efficiency of converting raw fuel into economically valuable products. (Basically, cogeneration won double duty from fuel.) An old process, being common in the United States early in the twentieth century, cogeneration lost favor, as utilities found they could exploit economies of scale and produce cheap, central station power for industrial customers. With the new need to improve energy efficiency in light of the energy crisis, President Carter wanted to see a resurgence of cogeneration technology.

Public Utility Regulatory Policies Act's Section 210 provided the means by which cogeneration would become more practical. For example, it eliminated the major impediment for cogenerating firms by requiring utility companies to purchase excess power from the nonutilities. Implemented by state regulatory authorities, this requirement gave industrial companies an incentive to install equipment that produced process steam and electricity. The firms would use some of the power themselves, but they could also sell surplus electricity (produced more efficiently than utilities) to power companies, which would then distribute it over their grid to customers. Previously, if nonutility companies wanted to sell excess electricity, utilities had no obligation to purchase it, and they often offered low rates to sellers. Now, PURPA required utilities to buy the electricity at a rate that equaled the utilities' cost of producing the power themselves.

This section of PURPA also contained similar incentives for companies and individuals that wanted to produce power from renewable sources of energy. These sources typically used the movement of water in rivers, the flow of air, and the effect of sunlight to produce electricity in ways that did not require the combustion of fossil fuels. Research and development on water turbines, wind turbines, and solar-electricity conversion technologies had been going on slowly for decades, and the 1973 energy crisis provided a spur. But provisions in PURPA gave these efforts more momentum because they created a guaranteed market for the electricity produced by these nontraditional generating technologies.

IMPLEMENTATION OF THE LAW

Implementation of PURPA became a job first of the Federal Energy Regulatory Commission (FERC). Under the terms of the law, FERC needed to hold hearings and codify parts of the law for the country's regulatory commissions to employ when dealing with the new, privileged nonutility generators, known as qualifying facilities (QFs). In most cases, FERC chose to interpret elements of "Section 210" in ways that benefited these companies. It relaxed administrative procedures so the nontraditional power firms would not be subject to the same government oversight as regulated utilities. Federal Energy Regulatory Commission also interpreted the law liberally so cogenerators and renewable energy generators earned relatively high rates for the power they produced. Overall, FERC sought to encourage these new firms with as many incentives as possible.[4]

After completion of this review process—as well as two U.S. Supreme Court decisions that maintained FERC's interpretations—PURPA found its way to the states, where public utility commissions defined how the law would be put into practice. In most cases, implementation of PURPA did not elicit much excitement. However, in California, the law became one tool in the state's efforts to obtain environmentally sound electricity supplies. Moreover, energy efficiency and renewable energy fit nicely into the political agenda of Governor Jerry Brown, who took office in 1975, and to several of the activist members of the California Public Utilities Commission (CPUC) appointed by the unconventional chief executive.

The CPUC sought to make it easy for QFs to deal with utilities by designing "standard offer" contracts that contained the relevant terms and conditions for the producers and buyers of power. To provide flexibility to the QFs, the CPUC designed several standard offers, with the interim Standard Offer 4 becoming popular after its issuance in 1983. This contract provided for per-kilowatt hour payments from utilities that rose over the first 10 years of a 15–30 year contract period. The rising prices were designed to keep track of the cost of energy, which analysts believed would continue increasing during the 1980s. Moreover, the long-term contracts offered certainty to the fledgling nonutility companies and allowed them to obtain financing more easily from investors. Because of these generous provisions—especially the escalating prices that contrasted the declining price of energy in 1984 and later—scores of nonutility generating companies signed up for the interim contracts, promising to build as much as 15,000 MW of capacity. Though the companies suffered no penalties if they failed to live up to their promises, regulators realized that they could see a glut of unused capacity if they allowed more companies to accept this offer. Consequently, the CPUC suspended it in 1985 and developed other standard offer contracts for nonutility developers. Nevertheless, those firms that had obtained the Standard Offer 4 contract could still take advantage of its lucrative terms.[5]

SPUR FOR TECHNOLOGICAL INNOVATION

Combined with tax incentives offered to small power producers by federal and state governments, PURPA accomplished its goal of spurring innovation in several technologies. Cogeneration technology had roots that went back to the early 1900s, but the law encouraged manufacturing companies to make improvements in it to exceed requirements to obtain at least 42.5%–45% thermal efficiencies. (Utility power plants typically converted only about 35% of raw fuel into electricity.) More important, perhaps, manufacturers exploited mass production techniques for several of the cogeneration plants' components, allowing construction of a plant for between \$800 and \$1200/kW in 1986. This capital cost compared favorably to utility costs for traditional fossil fuel and nuclear plants, which cost between about \$900 and \$5000/kW in 1984. And despite the lack of economies of scale, something that utilities prized in their huge plants, the cogenerating companies produced two salable products—electricity and heat—which enabled them to earn money while simultaneously meeting PURPA's goal of producing electricity in a more efficient fashion. Not surprisingly, cogeneration took off, making up 56% of all nonutility produced power in California in 1990 and 59% of nonutility power capacity nationwide in 1991. By 1992, almost 40,700 MW of the country's capacity came from cogeneration sources, up from 10,500 MW in 1979. In short, PURPA stimulated a new segment of the electric power business, as companies, such as Cogentrix, AES Corporation, and affiliates of Bechtel Power, General Electric, Mobil Oil, and Destech Energy, exploited the law's provisions.

Public Utility Regulatory Policies Act also stimulated the use of gas turbines used for making electricity. This technology had a long heritage, used for turning generators and producing electricity during peak usage times. But the small turbines (usually rated from 10 to 25 MW) had a reputation for poor reliability and low fuel efficiency. During the 1980s, however, research on gas turbines expanded, largely because of funding from the U.S. Department of Defense, which wanted smaller and more energy-efficient gas turbines for use in military aircraft. Manufacturers of these aeronautical turbines, such as General Electric, shrewdly realized that the technology could easily be transferred for stationary use in power plants, and they produced improved turbines for this application. To employ the turbines in the manner sanctioned by PURPA, manufacturers created systems that recovered energy produced from the gas engine's exhaust and reused it to heat water in a conventional steam turbine generator. These "combined cycle" units reached thermal efficiencies of $>40\%$ in the 1980s. Small and easy to install, these units also gained popularity in the mid-to-late 1980s, when the price of natural gas declined. At the end of 1992, nonutility companies employed gas turbines to provide almost 39% of their capacity. The cost of their delivered electricity ranged from about 3.2 to 5.5 cents/kWh, comparing favorably to the average cost of power produced by regulated power companies.

Beyond these more traditional technologies, wind turbine technology developed rapidly as a result of PURPA. Though used for pumping water and for producing small amounts of electricity on farms for decades, wind turbine technology had previously not been exploited to produce enough power for distribution by electric utilities. After the energy crisis, the federal government spent heavily to produce large wind turbines for utility use—some as large as 3.2 MW each—but the hardware failed. Much more successfully, entrepreneurs working under the impetus of PURPA fashioned smaller turbines (between 0.05 and 0.5 MW) to use in clusters, with the aggregated electricity sold to power companies. Some components of the small turbines could be mass produced or adapted from other applications, thus allowing the wind companies to pare costs. And because of the beneficial terms of California's standard offer contracts, many wind firms took root in the Golden State, even though wind resources may have been better in other parts of the country. Wind powered generation increased dramatically in California, rising from 50 million kWh produced in 1983 to 3268 million kWh generated in 1994.[6] Research and development in the technology continued, such that the technology in 2005 produces power at costs comparable to fossil fuel (including natural gas) generators, and more cheaply than any other nonhydro renewable resource.[7]

Technologies that employed the sun to make electricity also benefited from PURPA (and other laws, especially in California). Entrepreneurs working on solar photovoltaic cells, for example, designed and tested novel designs and could obtain some income because of PURPA's guarantee of a market for the cells' power. Largely because of the research, the cost of solar cell electricity declined from about 90 cents/kWh in 1980 to about 20 cents/kWh in 1995. Though still not commercially viable without additional subsidies, except in niche applications, photovoltaic technologies nevertheless improved dramatically. Meanwhile, another solar technology almost reached commercial success as a result of PURPA. Explicitly taking advantage of the law's provisions (and, again, other incentives offered in California) Luz International, a small entrepreneurial firm, built a system that employed parabolic mirrors to focus sunlight onto tubes containing oil. The heated oil transferred energy to water, which turned into vapor and powered a turbine-generator to produce electricity. Between 1984 and 1990, the company built several iterations of the system and produced power that dropped from 25 to 8 cents/kWh during that period. Overall, PURPA stimulated the development of novel power generation technologies, a fact noted as early as 1985 by analysts at the Congressional Office of Technology Assessment, which viewed the law as the motivator of several "first generation commercial applications" of technologies that may have great value in the future.[8]

PURPA AND DEREGULATION: UNINTENDED CONSEQUENCES OF THE LAW

Beyond its stimulation of research and development on small-scale technologies, PURPA unintentionally began the process of deregulating the American electric utility industry.[9] It did so largely in two ways. First, as noted earlier, PURPA eliminated the barrier to entry in the generation sector of the utility business. Under the regulatory framework that had existed since the beginning of the twentieth century, utility companies enjoyed status as regional monopolies that permitted them to generate, transmit, and distribute electricity without competition. But PURPA gave a special class of unregulated companies the right to produce power, effectively enabling them to compete with the formerly protected firms. In short, PURPA invalidated the notion that only integrated monopolies could effectively operate in the utility business.

Second, and perhaps more important, the success of some PURPA QFs in making a profit by selling power to utilities suggested that the regulated power companies no longer had a legitimate claim as natural monopolies. Going back to the late nineteenth century, the notion of natural monopoly helped justify the existence of regulation of utility companies in the first place. The academic and political principle held that in some industries (such as the railroad, streetcar, water delivery, and electric power industries), companies needed to invest heavily in technology and would only do so if they obtained guaranteed markets. Competition would also be detrimental to the public good because of the need for duplicate equipment, which raised costs, to serve the same customers. (Before regulation, streets of some cities were cluttered with wires from different companies, each seeking to serve the same lucrative customers.) Moreover, when several firms competed for a fixed number of customers, none could exploit the largest and best generation technologies that emerged in the early twentieth century. Steam-turbine generators, in particular, showed huge economies of scale. If two or more companies contested the same customer base, none could employ the biggest generating unit. But if only one firm served all customers, it could buy the largest equipment and reduce its unit cost for the power. Swaying politicians, this logic encouraged the creation of regulatory bodies that oversaw the natural monopoly utility companies, ensuring (in principle), that the benefits of the utilities' unusual status would flow to customers in the form of lower rates and good service.

The Public Utility Regulatory Policies Act challenged this rationale for regulation and, therefore, started discussion of further deregulation of the utility industry. It did so by spurring creation of nonutility companies that produced power at competitive prices. Even though the QFs did not exploit economies of scale like utilities did, they produced power with higher thermal efficiencies (in the case of cogenerators), and they sold the heat byproduct to gain economic advantages over utilities. They, therefore, often could beat utility costs by a margin of 5%–15% in the mid-1990s. This fact was noted by some politicians and regulators, who argued that the financial success of some QFs indicated that utilities no longer constituted natural monopolies. After all, natural monopolies supposedly produced power more efficiently and more cheaply than could competitors. But some PURPA QFs demonstrated that they could produce lower priced electricity than could the protected utilities. Following this logic further, the questioning of the natural monopoly rationale led to challenges to the notion of regulation. Congressmen, federal regulators, and some utility leaders themselves started asking why regulation should exist when utility monopolies no longer could be justified as "natural" anymore.

This questioning of regulation in the utility sector did not occur in a vacuum. Throughout the 1970s and 1980s, economists and politicians in the United States and elsewhere began challenging the value of regulation of various industries as the best way to provide services to customers. Starting with President Carter's efforts, the airline industry began the process of deregulation. Similar moves to introduce market forces occurred in the natural gas, petroleum, railroad freight transportation, and financial services industries, only to be followed (during President Reagan's tenure in office) by deregulation of the telecommunications industry. Similar deregulation of various industries occurred in England and several other countries, yielding (in many cases) declining prices, improved services, and technological innovation in industries that had previously appeared stagnant.

With the questioning of natural monopoly status in the power industry, stimulated partly by the experience of PURPA, and the apparent success of deregulation in other businesses, many people called for restructuring of the electricity infrastructure. President George H.W. Bush heeded these calls and proposed, after the Gulf War of 1991, to employ market forces to increase domestic fuel production and to improve the efficiency of energy use. One provision of the Energy Policy Act of 1992 gave states the right to allow their transmission network to serve as a common carrier so any electricity producer could sell power to any customer. Essentially, the law enabled states to begin competition on the retail level. Taking advantage of the legislation, several states in the late 1990s established competitive retail frameworks. By September 2001, 23 states (and the District of Columbia) had passed restructuring laws, while other states used the regulatory process to reduce their oversight and introduce more market forces into the industry.[10]

CONCLUSION

While PURPA contributed to the deregulation of the power industry, that same deregulation also minimized the importance of PURPA. Though the law remains on the books, despite the best efforts of repeal proponents,[11] who think the law has led to overpriced and unneeded power, terms of the Energy Policy Acts of 1992 and 2005 make some elements of PURPA less significant. For example, the 1992 law created a class of independent power producers known as exempt wholesale generators. While these generators do not receive guaranteed payments from utilities, as do QFs, they also are not bound by PURPA's requirements for overall energy efficiency. Entrepreneurial companies have taken advantage of this feature of the new law and have built scores of plants that sell power to an open market of customers. They have eclipsed QFs as the primary nonutility providers of power to the nation's electricity grid.[12]

Nevertheless, the significance of PURPA should not be overlooked just because the law no longer has as much practical meaning as it did in the 1980s. Intended largely to reform the way utility companies charged customers for electricity, the statute unexpectedly had at least two major consequences. First, a benign-looking part of the law provided the financial incentive that motivated entrepreneurs to perform research and development activities on small-scale renewable energy technologies, such as wind turbines and solar energy systems. Combined with tax breaks offered by some states, these technologies saw rapid improvement and the ability to produce electricity at greatly reduced costs. Second, the law helped undermine the rationale for utilities' status as regulated natural monopolies. It did so by encouraging firms to improve cogeneration and gas-turbine technologies and to make them commercially feasible in small sizes. Owners of the PURPA-induced qualifying facilities employed these technologies to break down the barriers to entry in the utility business by allowing them to compete with utility companies in the generation sector. In the process, they challenged the notion that utilities deserved recognition as special noncompetitive enterprises and that they required oversight by government regulatory bodies. Within a political environment that valued the deregulation of industries, PURPA's questioning of regulation motivated, to a large extent, the restructuring of the utility industry. That restructuring process continues into the early part of the twenty-first century.

REFERENCES

1. This article draws heavily from Hirsh, R.F. *Power Loss: The Origins of Deregulation and Restructuring in the American Electric Utility System*; MIT Press: Cambridge, MA, 1999: Hirsh, R.F. PURPA: *The spur to competition and utility restructuring Elec. J.* **1999**, *12* (7), 60–72.
2. Carter, J. President's proposed energy policy. Vital Speeches of the Day **1977**, *43* (14), 418–420.
3. Public Utility Regulatory Policies Act, Public Law 95–617, signed 9 November 1978.
4. Federal Energy Regulatory Commission Small power production and cogeneration facilities; regulations implementing Section 210 of the public utility regulatory policies act of 1978. Fed. Regist. **1980**, *45* (25), 12–215: FERC Power production and cogeneration facilities—qualifying status. Fed. Regist. **1980**, March (20), 17–959.
5. Ahern, W.R. Implementing avoided cost pricing for alternative electricity generators in California. In *New Regulatory and Management Strategies in a Changing Market Environment*; Trebing, H.M., Mann, P.C., Eds., MSU Public Utilities Papers: East Lansing, MI, 1987; 408.
6. California Energy Commission, California Statistical Abstract 1996 http://www.dof.ca.gov/html/fs_data/STAT-ABS/sec_J.htm, obtained 18 July 1997.
7. Roth, I.F.; Ambs, L.L. Incorporating externalities into a full cost approach to electric power generation life-cycle costing. Energy **2004**, *29*, 2125–2144.
8. The previous paragraphs rely on information in Hirsh, *Power Loss*, 101–117. Also see Guey-Lee, L. Renewable electricity purchases: history and recent developments. In *Renewable Energy 1998: Issues and Trends*, U.S. Department of Energy, Energy Information Administration, at http://www.eia.doe.gov/cneaf/solar.renewables/rea_issues/renewelec_art.pdf, obtained 12 April 2005.
9. Strictly speaking, one should really be speak about "restructuring", rather than "deregulating", the electric utility system. Deregulation generally means the end of all government oversight and the existence of a totally free market. While the generation sector of the utility business became somewhat competitive under PURPA, it retained a large amount of regulation that oversaw the interactions between utilities and qualifying facilities. More accurately, restructuring describes events that maintain a government presence in all utility sectors to ensure that (in theory) the new wholesale and retail markets operate efficiently and in the public interest.
10. U.S. Department of Energy, Energy Information Administration, at http://www.eia.doe.gov/cneaf/electricity/chg_str/regmap.html, obtained 15 October 2001.
11. Salpukas, A. Utilities See Costly Time Warp In '78 Law. In *New York Times*, Vol. 5 1994. Section D; 1. See efforts of the PURPA Reform Group, an ad hoc organization made up of utility companies that objected to allegedly higher costs caused by implementation of the law. Adelberg, A. Way off on PURPA. Elec. J. **1999**, *12* (10), 2–3. Mr Adelberg was co-chair of the PURPA Reform Group.
12. U.S. Department of Energy, Energy Information Administration *Electric Power Annual 2003*. See 2, 8, 13 [OE/EIA-0348(2003)].

Pumped Storage Hydroelectricity

Jill S. Tietjen
Technically Speaking, Inc., Greenwood Village, Colorado, U.S.A.

Abstract
Pumped storage hydro is a form of hydroelectric power generation for electric utilities that incorporates an energy storage feature. The fuel, water, moves between two reservoirs—an upper and a lower—with a significant vertical distance between them. Water is stored in the upper reservoir until such time as the utility determines that it is economic to use the water to produce electricity for the system, usually to keep coal-fired and nuclear power plants operating at economic levels during low load periods, such as at night and on weekends. Pumped storage is the most widespread energy storage system in use on power networks and is used for energy management, frequency control, and provision of reserves. Efficiency of any specific pumped storage facility, which is primarily dependent on the height between the upper and lower reservoirs, ranges from 70 to 85%.

INTRODUCTION

Pumped storage hydro is a form of hydroelectric power generation for electric utilities that incorporates an energy storage feature.[1] The fuel, water, moves between two reservoirs—an upper and a lower—with a significant vertical separation (see Fig. 1). Water is stored in the upper reservoir until such time as the utility determines that it is economic to use the water to produce electricity for the system. The water in the upper reservoir is stored gravitational energy.[2] When the water is released, the force of that water spins the blades of a turbine that connect to a generator, which produces electricity.[3]

Water is generally released from the upper reservoir to produce electricity during the daylight or on-peak hours. After passing through the turbines, the water is discharged into the lower reservoir. At night and on weekends, or during light-load or off-peak hours, water is pumped from the lower reservoir into the upper reservoir by the turbines, which have been reversed to work as electric motor-driven pumps.

In the upper reservoir, the water is essentially stored energy. Water can be stored for a long or short time in the upper reservoir, depending on the needs of the utility. A vertical separation of at least 100 m (328 ft) is necessary to make a pumped hydro facility economic.[4] The height difference between the upper and lower reservoirs is called the head. The amount of potential energy in the water is directly proportional to the head, so the greater the height, the more energy that can be stored.[1,5]

Pumped storage hydro was first used in Italy and Switzerland in the 1890s. In 1929, the first major pumped storage hydroelectric plant in the United States, Rocky River, was built in New Milford, Connecticut. By 1933, reversible pump-turbines with motor generators had become available. The turbines could operate as both turbine-generators and in reverse, as electric motor-driven pumps.[6,7]

About 3% of total global generation capacity—more than 90 GW—is pumped storage capacity. In 2000, the United States had 19,500 MW of pumped storage facilities in operation.

The largest pumped storage facility in the United States is in Bath County, Virginia, which has a capacity of 2,100 MW. Pumped storage plants are characterized by long construction times and high capital expenditures.[6,8,9]

Pumped storage is the most widespread energy storage system in use on power networks and is used for energy management, frequency control, and provision of reserves. Efficiency of any specific pumped storage facility, which is primarily dependent on the height between the upper and lower reservoirs, ranges from 70 to 85%.

Environmental issues associated with the construction of pumped storage plants often parallel those for conventional hydroelectric plants and range from water-resource and biological effects to potential damage to archaeological, cultural, and historical sites.[6]

HYDROELECTRIC GENERATION

In conventional hydroelectric generation, hydraulic turbines rotate due to the force of moving water (its kinetic energy) as it flows from a higher to a lower elevation. This water can be flowing naturally in streams or rivers, or it can be contained in manmade facilities such as canals, reservoirs, and pipelines. Dams raise the water level of a stream or river to a height sufficient to create an adequate head (height differential) for electricity generation.[10]

Keywords: Energy storage; Hydroelectric; Reservoir; Pumped storage; Reversible pump-turbine; Francis turbine; Turbine-generator; Head; Flow.

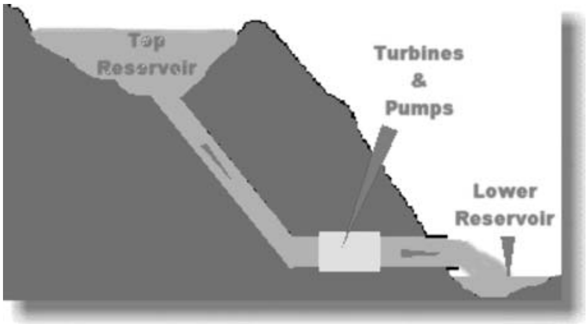


Fig. 1 Pumped storage hydro configuration.

If the dam stops the flow of the river, water pools behind the dam to form a reservoir or artificial lake. As hydroelectric generation is needed by the electric utility, the water is released to flow through the dam and powerhouse. In other cases, the dam is simply built across the river, and the water moves through the power plant or powerhouse inside the dam on its way downstream.[11]

In either case, as the water actually moves through the dam, the water pushes against the blades of a turbine, causing the blades to turn (see Fig. 2). The turbine converts the energy in the form of falling water to rotating shaft power to turn a generator, which produces electricity.[11,12]

The selection of the best turbine for any specific hydroelectric site is primarily dependent on the head (the vertical distance through which the water falls) and the water flow (measured as volume per unit of time) available. Generally, a high-head plant needs less water flow than a low-head plant to produce the same amount of electricity.[12,13]

The power available in a stream of water is

$$P = \eta \rho g h V$$

where: P, power (J/s or watts); η, turbine efficiency; ρ, density of water (kg/m^3); g, acceleration of gravity (9.81 m/s^2); h, head (m; the difference in height between the inlet and outlet water surfaces); V, flow rate (m^3/s).[14]

For SI units, this equation can be roughly approximated as

$$\text{POWER(kW)} = 5.9 \times \text{FLOW} \times \text{HEAD}$$

where FLOW is measured in cubic meters per second and HEAD is measured in meters.[15] In general terms, for IP units, 1 gal/s falling 100 ft can generate 1 kW of electrical power.[16]

Hydroelectric power is generally found in mountainous areas where there are lakes and reservoirs and along rivers. Hydroelectric power currently provides about 10% of all the electricity produced in the United States. Hydroelectricity provides about one-fifth of the world's electricity. Worldwide capacity is 650,000 MW.[11,16]

Hydroelectric power is characterized as a renewable resource. The fuel, water, is replenished by rain and snowfall. With an existing plant, it is the cheapest way today to generate electricity. Producing electricity from hydropower is so economical because when the dam is built and the equipment has been installed, the flowing water has no cost. In addition, the dams are very robust structures, and the equipment is relatively simple mechanically. Hydro plants are dependable and long lived, and their maintenance costs are low compared with those of most other forms of electricity generation, including fossil-fired and nuclear generation. Many hydroelectric plants do not even require staff, further contributing to their low ongoing maintenance costs.[13]

Pumped storage hydro generation is a specific kind of hydroelectric generation requiring an upper and lower reservoir, and special equipment that can both generate power as water flows downhill and then reverse and serve as a pump to move the water back uphill.

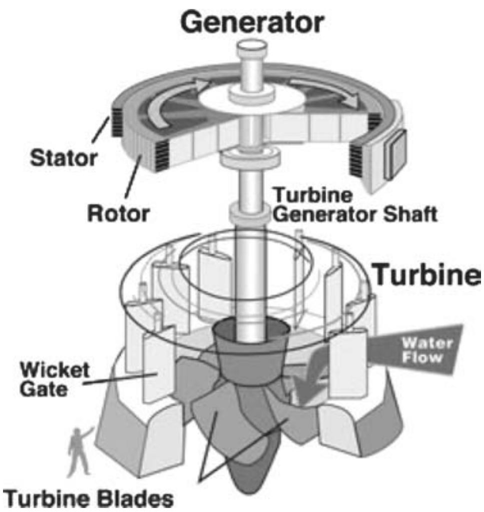


Fig. 2 Hydroelectric turbine-generator.

FACILITY DESCRIPTION

A pumped storage hydro power plant typically has three major components: the upper reservoir, the lower reservoir, and the pumping/generating facilities. The pumped storage facility represented in Fig. 3 is Duke Power Company's Jocassee plant in South Carolina.

The water from the upper reservoir used for electricity generation is allowed into the intake shaft through the opening of the headgates. Water moves through the high-pressure shaft and steel-lined power tunnel until it reaches the turbines in the powerhouse.[1] The water turns the turbines, which then drive the generators to produce electricity. Then the water moves through the tailrace tunnel until it is discharged into the lower reservoir, which in Fig. 3 is Lake Jocassee. Water discharge capacity in the upper reservoir can require several hours to several days.[6]

Pumped storage facilities can be categorized as "pure" or "combined." Pure pumped storage plants, also referred to as modular pumped storage or MPS, continually shift water between an upper and a lower reservoir. Combined pumped storage plants also generate their own electricity like conventional hydroelectric plants through natural stream flow. Modular pumped storage systems tend to be much smaller than conventional hydroelectric power plants. They use closed water systems that are artificially created instead of natural waterways or watersheds. The water for MPS usually is put into the system only when it begins operation, either from groundwater or possibly from municipal wastewater.[8,10]

When water is to be pumped back into the upper reservoir of a pumped storage plant, it flows from the lower reservoir into the powerhouse, where reversible pump-turbines (usually, a Francis turbine design) pump it back into the upper reservoir. The hydraulic units are designed to operate as pumps when rotating in one direction and as turbines when rotating in the opposite direction. Similarly, motors that drive the pumps can be reversed to act as generators. Fig. 4 contains a schematic diagram showing the directions of water flows and rotation of the pump-turbine for each mode of operation.[8,9]

Many turbines used at pumped storage plants are Francis turbines, named for American engineer James B. Francis (1815–1892), who worked to enhance turbine design. Francis turbines are capable of applications of 2–800 m (approximately 6.5–2,600 ft). The turbines are individually designed for each site to operate as efficiently as possible and to match each site's flow conditions. These turbines may be designed for a wide range of heads and water flows. Francis turbines are very expensive to design, manufacture, and install, but they operate for decades.[14,17–19]

The efficiency of pumped storage plants generally ranges from 70 to 85%. This means that 70%–85% of the electrical energy used to pump the water into the upper reservoir is actually generated when water flows back down through the turbines to the lower reservoir. The losses of energy occur due to evaporation from the exposed water surface (in both the upper and lower reservoirs) and to the mechanical efficiency losses during conversion from flowing water to electricity. In hot climates or windy areas, the evaporation losses will be higher. Similarly, mechanical losses are higher with older equipment.[8]

HOW PUMPED STORAGE IS USED BY UTILITIES

Pumped storage hydro serves an important function in the portfolio of resources available to utilities. They provide the most significant and most cost-effective means of storage of energy on a scale useful for a utility. It is not usually feasible and/or economic for utilities to turn down or reduce load on large nuclear and coal-fired (so-called thermal) units during low load periods—generally, at night. Using electricity from these thermal units to provide the power needed to pump water into the upper reservoir allows the utility to have the water to use during higher load periods the next day to generate electricity through the turbines of the pumped storage power plant and to keep their coal and nuclear units operating at more efficient levels.[5]

Pumped storage hydro is economical because it flattens the variations in load on the power grid, permitting thermal plants (coal-fired and nuclear) that provide electricity to continue operating at their most efficient capacity while reducing the need to build special power plants that will run only at peak demand, using more costly generation methods.[8]

Fig. 5 shows a typical pattern for pumped storage usage by a utility. Here, power is generated by the pumped storage facility during the higher load hours of 7:00 A.M. to 10:00 P.M., and pumping (which is usage of electricity from the grid) occurs in the off-peak hours.

In addition to energy management, pumped storage systems are important components in controlling electrical network frequency and in providing reserve generation. Thermal plants are much less able to respond to sudden

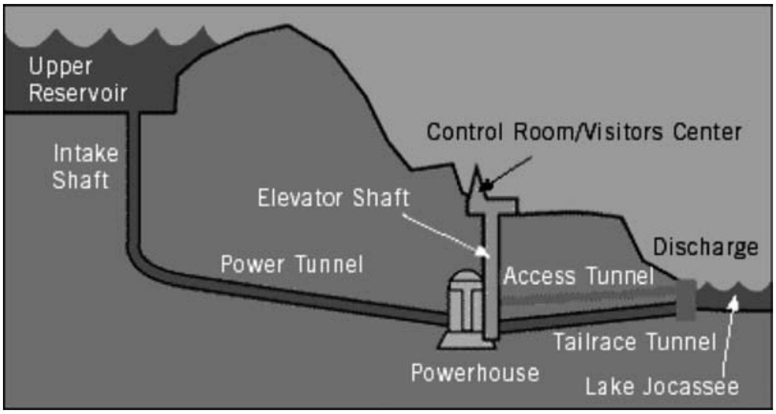


Fig. 3 Pumped storage power plant.

Pum–Ren

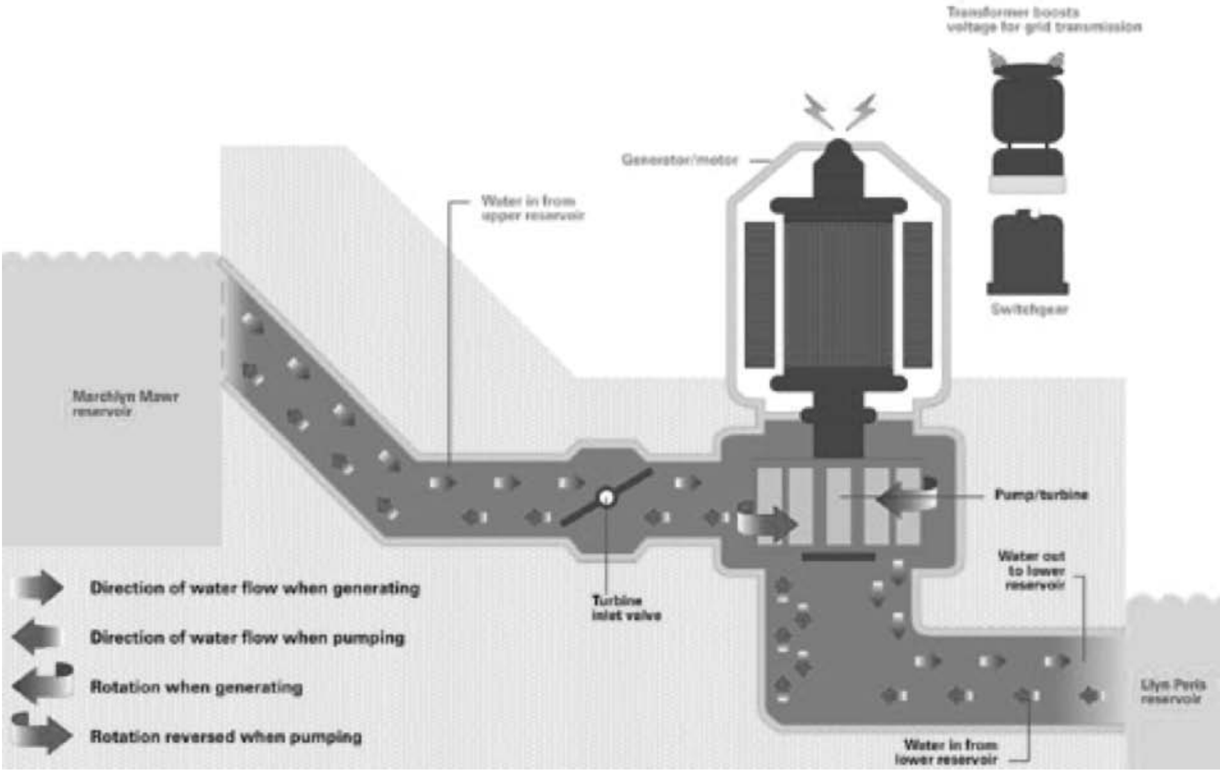


Fig. 4 Schematic of pumped storage hydro operation.

changes in electrical demand, which cause frequency and voltage instability. Pumped storage plants can respond to those changes in seconds. This is particularly desirable in the case of a unit's becoming unavailable or forced out of service or on utility systems with high amounts of intermittent resources, such as solar and wind. The Dinorwig pumped storage facility in north Wales, United Kingdom, for example, can go from 0 MW to full capacity of 1,320 MW in 12 s and usually can stay at this level until other generating units on the utility's system can be brought online.[1,4,8]

COST AND ECONOMICS

The cost effectiveness of a pumped storage hydroelectric facility depends on the topography of an area and on the types and sizes of power plants in the electric utility's system. The capital cost of the facility will be dependent on the height of the head and the geography in the area. Adequate resources need to be available consistently to provide the pumping energy required economically.

Higher head ranges, varying between 300 and 600 m (approximately 1,000–2,000 ft), and relatively steep topography are generally desirable for a cost-effective facility to be designed, built, and operated. Two main parameters affect the costs of pumped storage facilities: the ratio of waterway length to head (l/h) and the overall head of the project. A low l/h ratio will result in shorter water passages and will reduce the need for surge tanks to control transient flow conditions. Higher head projects require smaller volumes of water to provide the same level of energy storage and smaller waterway passages for the same level of power generation. Most desirable pumped storage sites have heads in excess of 300 m (approximately 1,000 ft) with l/h ratio of 6 or less.[20]

New pumped storage projects have not been built in the United States for quite a number of years. In 2006, the Lake Elsinore Advanced Pumped Storage Project is being proposed in California and is under review by the Federal Energy Regulatory Commission. Two separate alternatives have been proposed, including pumped storage and pumped storage plus a conventional hydroelectric facility, at costs ranging from $720 million to $1.3 billion, including the associated required transmission for 500–1,275 MW.[21]

The cost effectiveness of the pumped storage power plant during daily operation for an electric utility depends on the cost of the power used to pump the water back uphill, the efficiency of the unit, and the cost of the power at the margin during the utility's on-peak period. Consider the example of the following hypothetical utility (Utility A), whose level of so-called production costs—costs

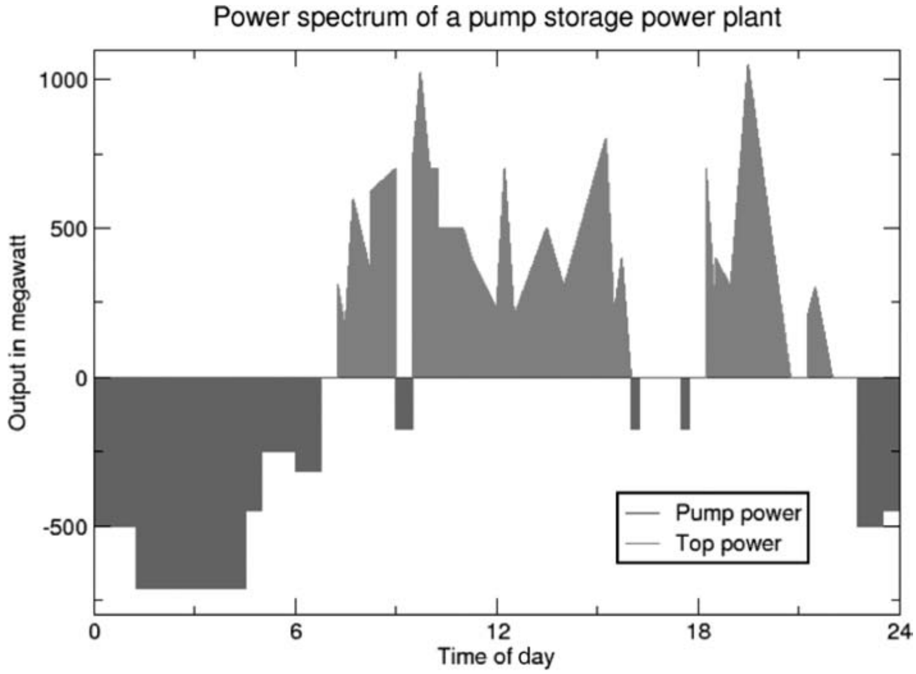


Fig. 5 Power spectrum of a pumped storage power plant.

including only fuel and operating and maintenance (O&M) costs—are as shown.

Utility A has a peak load of 10,000 MW. Utility A has three 750 MW nuclear power units that generate electricity at a production cost of 1.7 cents/kWh. Additional generation resources include a variety of coal-fired generating units totaling 7,000 MW that have an average production cost of 2.5 cents/kWh. Utility A also has 1,250 MW of natural gas-fired combustion turbines that produce power at an average price of 5.5 cents/kWh. There is also 1,000 MW of conventional hydroelectric capacity on Utility A's system.[22]

The following table summarizes the resources and the costs for Utility A. Utility A also has a 750 MW pumped storage facility that is 70% efficient (Table 1).

Because the cost of using the pumping energy is significantly lower than the production cost of using natural gas, Utility A can cost-effectively use the nuclear and coal resources at night to pump water back uphill and then use generation from the pumped hydro facility during the day instead of running its natural gas-fired combustion turbine units.

ENVIRONMENTAL ISSUES

Issues related to permitting of conventional hydroelectric and on-stream pumped storage hydroelectric facilities include

- Water-resource impacts: stream flows, reservoir surface area, groundwater recharge, water temperature, turbidity, and oxygen content
- Biological impacts: displacement of terrestrial habitat, alteration of fish migration patterns, and other impacts due to changes in water quality and quantity
- Potential damage to archaeological, cultural, or historic sites
- Visual-quality changes

Table 1 Resources and costs for Utility A

Generation type	MW	Production cost (cents/kWh)	Cost of energy from pumped hydro using this capacity to pump
Conventional hydroelectric	1,000	0	Already used to serve load, not available to use for pumping
Nuclear	2,250	1.7	2.21
Coal	7,000	2.5	3.25
Natural gas	1,250	5.5	Too expensive to use this fuel to pump

- Loss of scenic or wilderness resources
- Increased risks of landslides and erosion
- Gain in recreational resources

These concerns are much less significant for MPS plants, because MPS operate in a closed loop and are not associated with natural waterways and watersheds. Usually, MPS are specifically not located near existing rivers, lakes, streams, and other sensitive environmental areas to avoid the regulatory lag time and complexity associated with combined pumped storage hydroelectric facilities.[23]

CONCLUSION

Pumped storage hydroelectric facilities offer many benefits to utilities, including energy management, frequency control, and the provision of reserves. Utilizing water as a fuel, these facilities now provide about 3% of the world's generating capacity. These plants are the most economical in large utility systems, where the pumping energy can come from large coal-fired and nuclear units, and can keep those units from having to reduce load at night and on weekends.

REFERENCES

1. Andy Darvill, Energy Resources. Pumped Storage Reservoirs—storing energy to cope with big demands. Available at: http://home.clara.net/darvill/altenerg/pumped.htm (accessed March 2006).
2. U.S. Department of Energy—Energy Efficiency and Renewable Energy Distributed Energy Program. Pumped Hydro. Available at: www.eere.energy.gov/de/pumped_hydro.html (accessed January 2006).
3. Tennessee Valley Authority, Hydroelectric Power. Available at: www.tva.gov/power/hydro.htm (accessed March 2006).
4. University of Alaska, Fairbanks. Available at: www.uaf.edu/energyin/webpage/webpagewithframes/pumpedhydro.text.htm (accessed January 2006).
5. Hydroelectricity. From Wikipedia. Available at: http://en.wikipedia.org/wiki/Hydroelectricity (accessed March 2006).
6. Electricity Storage Association. Technologies & Applications, Technologies, Pumped Hydro Storage. Available at: www.electricitystorage.org/tech/technologies_technologies_pumpedhydro.htm (accessed March 2006).
7. U.S. Bureau of Reclamation. The History of Hydropower Development in the United States. Available at: www.usbr.gov/power/edu/history.html (accessed April 2006).
8. Pumped-Storage Hydroelectricity. From wikipedia. Available at: http://en.wikipedia.org/wiki/Pumped-storage_hydroelectricity (accessed March 2006).
9. Encyclopedia Brittanica. Waterpower: Pumped Hydro Storage. Available at: www.brittanica.com/ebi/article-210089 (accessed March 2006).
10. California Energy Commission. Hydroelectric Power in California. Available at: www.energy.ca.gov/electricity/hydro.html (accessed April 2006).
11. California Energy Commission. Energy Story: Chapter 12—Hydro Power. Available at: www.energyquest.ca.gov/story/chapter12.html (accessed April 2006).
12. Micro Hydropower Basics: Turbines. Available at: www.microhydropower.net/turbines.html (accessed April 2006).
13. Anderson High School NEED Team, Learn About Energy.org. Part 11, Hydropower. Available at: www.learnaboutenergy.org/focus/part11.htm (accessed April 2006).
14. Water Turbine. From wikipedia. Available at: http://en.wikipedia.org/wiki/Water_turbine (accessed April 2006).
15. ICLEI, Stuart Baird. Hydro-Electric Power. Available at: www.ecology.com/archived-links/hydroelectric-energy (accessed April 2006).
16. Union of Concerned Scientists. How Hydroelectric Energy Works. Available at: www.ucsusa.org/clean_energy/renewable_energy_basics/how-hydroelectric-energy-works_html (accessed April 2006).
17. James B. Francis. From wikipedia. Available at: http://en.wikipedia.org/wiki/James_B_Francis (accessed April 2006).
18. Francis turbine. From wikipedia Available at: http://en.wikipedia.org/wiki/Francis_turbine (accessed March 2006).
19. The Worlds of David Darling, The Encyclopedia of Alternative Energy and Sustainable Living. Francis turbine. Available at: www.daviddarling.info/encyclopedia/F/AE_Francis_turbine.html (accessed April 2006).
20. World Bank. Pumped-Storage Projects. Available at: www.worldbank.org/html/fpd/em/hydro/psp.stm (accessed March 2006).
21. Jose Carvajal. High Hydro Plant Price Tag Not Seen as a Dealbreaker. Available at: www.nctimes.com/articles/2006/03/09/news/californian/21_21_563_8_06.txt (accessed April 2006).
22. Nuclear Energy Institute. Comparative Measures of Power Plant Efficiency. Available at: www.nei.org/index.asp?catnum=2?catid=262 (accessed April 2006).
23. Claifornia Energy Commission. Hydroelectric Power in California. Available at: www.energy.ca.gov/electricity/hydro.html (accessed April 2006).

Pumps and Fans

L. J. Grobler

T. G. Vorster
School for Mechanical Engineering, North-West University, Potchefstroom, South Africa

Pum–Ren

Abstract
Pumps and fans are used in the commercial and industrial sectors to transfer fluids and gases and to consume a significant amount of energy. This entry focuses on energy management and cost saving opportunities that exist with pumping and fan systems. It gives an overview on the types of pumps and fans, their operating characteristics, theory, and power requirements, and highlights typical energy management opportunities through maintenance and retrofitting of pumping and fan systems.

INTRODUCTION

Pumps and fans are two of the most common pieces of equipment in use today in commercial and industrial applications. As such, they make up a large part of the energy used in these applications and therefore offer large scope for energy reduction measures.

Fans and pumps react within the same manner; therefore, the governing laws are similar for both fans and pumps. The difference between fans and pumps is in the fluid being moved; in the case of a fan, it is some type of gas, while the pump moves some type of liquid. A second, but less clear, distinction is that pumps are usually coupled directly to the drive motor while the fan uses a system of pulleys.

The aim of this article is to provide the reader with the basics of pump and fan system design, while highlighting the importance of proper design in energy usage reduction. The article also aims to highlight other energy reduction methods in use or which are currently being developed.

PUMPING SYSTEMS

The Internet Free Dictionary describes a pump as "a machine or device for raising, compressing, or transferring fluids."

This is achieved through the addition of energy to the fluid by various methods. These methods and the theory governing pumps and the types of pumps currently in use are discussed below.

Keywords: Pumping systems; Fan systems; Types of fans and pumps; Operating characteristics; Theory; Power requirements; Energy management opportunities; Retrofit opportunities.

Types of Pumps

There are two main categories of pumps that are identified by their basic principle of operation, namely positive displacement pumps and centrifugal pumps. Under these two basic types, there various subtypes. These are shown in Fig. 1.

The Centrifugal Pump

The term Centrifugal pump is a general description for machines that displace fluid through the addition of momentum provided by a spinning impeller. These pumps are commonly used in a wide range of industrial and commercial applications.

The basic centrifugal pump, known as a radial flow pump, as shown in Fig. 2, has rotational vanes that move the liquid away from the center of the impeller into the scrolled casing. A part of the kinetic energy imparted to the fluid is converted into pressure, which forces the liquid out of the discharge.

The other type of centrifugal pump is the axial flow centrifugal pump, which imparts energy to a fluid through a lifting motion by a set of propeller-shaped vanes. This motion results in an axial discharge.

A mixed flow centrifugal pump is a pump that uses a combination of radial and axial action.

The Positive Displacement Pump

The positive displacement pump works by trapping liquid in a cavity and then displacing it by some or other method to the pump discharge. These types of pumps provide an almost constant flow rate for a particular pump speed, which is independent of pressure differences and liquid characteristics. A gear type positive displacement pumps is shown in Fig. 3.

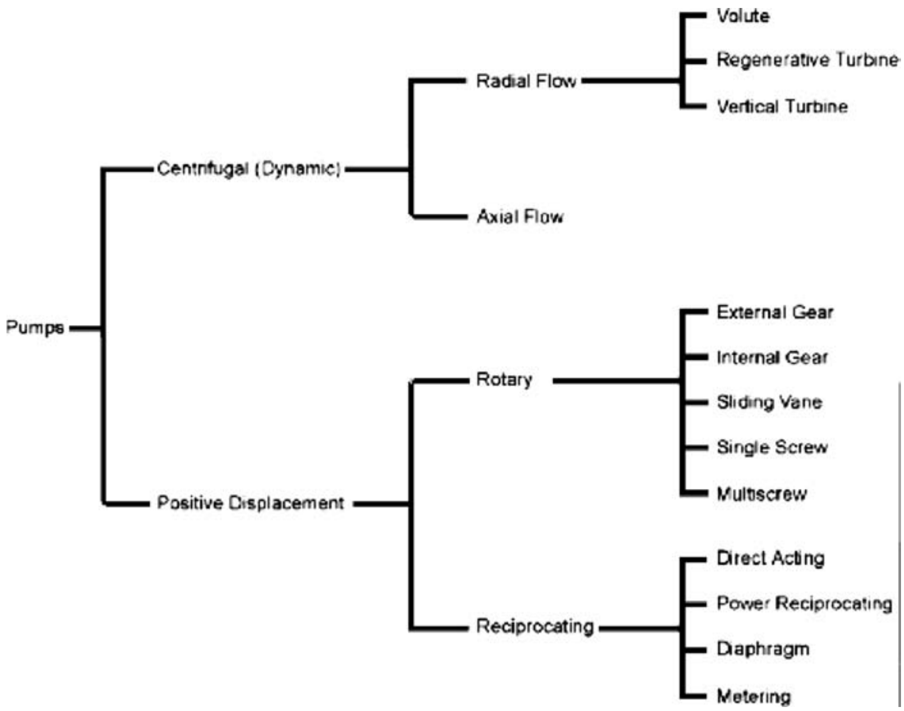


Fig. 1 Classification of pump types. Source: From ENEWS (see Ref. 1).

Pressure reducing devices are normally installed at the outlet of positive displacement pumps to limit the total head to that of the rated pump output head. Data for these types of pumps are usually found in the form of a table listing the flow rate, total head, and power. As positive displacement pumps are usually used to pump liquids other than water, the assistance of an expert may be needed to select a pump for a particular application.

Pump Theory

Pump Operating Characteristics

Pump performance is defined by the flow rate and head. Head is the pressure required to support a certain height of fluid. Head is expressed in "meters of water."

To explain this, visualize a $1 \times 1 \times 1\ m^3$ vessel containing water. The weight of the water contained in the vessel is 1000 kg. The pressure exerted on the bottom of the vessel is 1000 kg/m^2 which equates to 10 kPa. This is the equivalent pressure exerted by a 1 m column of water. This pressure is dependent on the specific gravity of the fluid, so if the vessel was to be filled with another liquid, the head would still be 1 m, but the equivalent pressure would change.

The relationship between equivalent pressure and head can now be equated as:

$$H = \frac{P}{9.81 \times SG} \quad (1)$$

where H, head of fluid (m); P, equivalent pressure (kPa); and SG, specific gravity of fluid (kg/m^3).

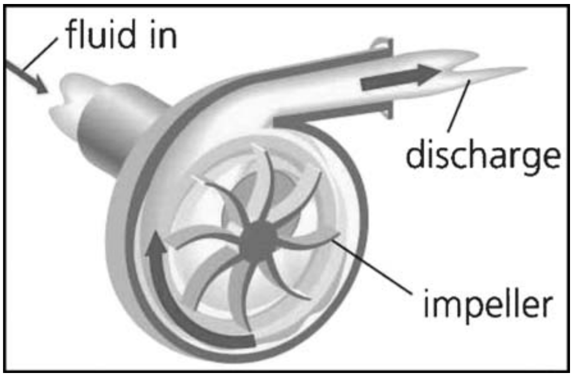


Fig. 2 A basic centrifugal pump. Source: From The Free Dictionary.

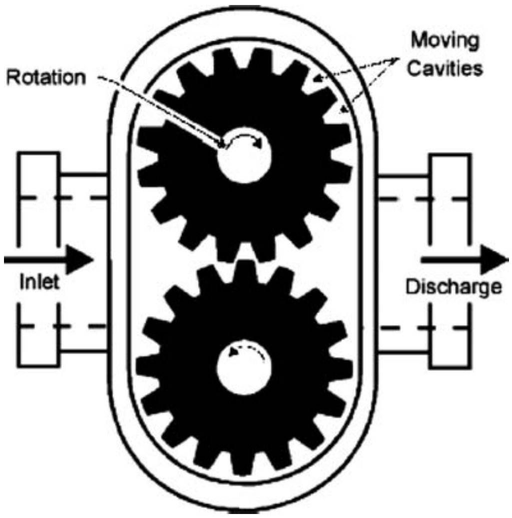


Fig. 3 A basic positive displacement pump.
Source: From ENEWS (see Ref. 1).

For water, this roughly equates to 1 m = 9.81 kPa.

The required head differential across a pump to move a liquid is called the total dynamic head. The output head or dynamic head of a pump must be equal or greater than the sum of the static head and friction head.

Friction Head. The friction and inertia losses, due to the resistance to the flow of a fluid through a pipe, is known as the friction head. These losses depend on the diameter, material, and length of pipe, as well as the type and amount of fittings used in the piping system. Friction head is proportional to the square of the flow rate, thus friction losses are zero if there is no flow.

Velocity Head. Velocity head is the amount of kinetic energy contained within a fluid for a specific flow rate. The relationship between the velocity head and fluid velocity is given in Eq. 2:

$$P_v = \frac{V^2}{2g} \tag{2}$$

where P_v, velocity head (m); v, fluid velocity (m/s); and g, gravity (9.81 m/s^2).

Because of velocity head is usually small for typical systems, it is incorporated in the friction head, but for gravity and low-pressure applications, it should always be considered.

Total Static Head. The head across a pump when there is no flow is called the total static head. Total static head is the sum of the elevation head and system pressure differential. Elevation head is the difference in height or elevation between the pump and the discharge of the system. The system pressure differential is the difference between the static suction head (the static head at the pump intake) and the static discharge head (static pressure at pump outlet).

Centrifugal Pump Theory

Centrifugal pump performance is defined by the impeller diameter, pump speed, flow rate, head, power, and fluid characteristics. Manufacturers provide data in graph form for each pump type, pump size, and pump. These graphs are usually based on water as the pumped fluid. A typical example of such a graph is shown below in Fig. 4.

The set of graphs in Fig. 4 show the head, power, and efficiency of the pump for a certain flow rate. Each of the numbered curves represents a different pump speed. Some pump curves incorporate different impeller diameters into the curves instead of pump speed. Most manufacturers recommend that only the middle section of the curves be used, as the outside sections may prove unstable.

The variables determining centrifugal pump performance are related by the Affinity Laws, which are used to develop the pump curves. These laws are given in Eqs. 3–9.

D and δ constant:

$$Q_1 = Q_2\left(\frac{N_1}{N_2}\right) \tag{3}$$

$$P_1 = P_2\left(\frac{N_1}{N_2}\right)^2 \tag{4}$$

$$P_1 = P_2\left(\frac{N_1}{N_2}\right)^2 \tag{5}$$

N and δ constant:

$$Q_1 = Q_2\left(\frac{D_1}{D_2}\right) \tag{6}$$

$$P_1 = P_2\left(\frac{D_1}{D_2}\right)^2 \tag{7}$$

$$W_1 = W_2\left(\frac{D_1}{D_2}\right)^3 \tag{8}$$

N and D constant:

$$W_1 = W_2\left(\frac{\delta_1}{\delta_2}\right) \tag{9}$$

where D, impeller diameter (m); Q, flow rate (m^3/h); P, total head; W, power (kW); N, pump speed (RPM); and δ, fluid density (kg/m^3).

If conditions in a system change, but the performance curves remain unchanged, the laws can be used to determine the new performance of the system. They may, however, not be used where changes in the layout of the piping causes changes in the system performance curve. These laws may also not be used to analyze systems with a significant static head component, unless the change to the system is less than 10%.

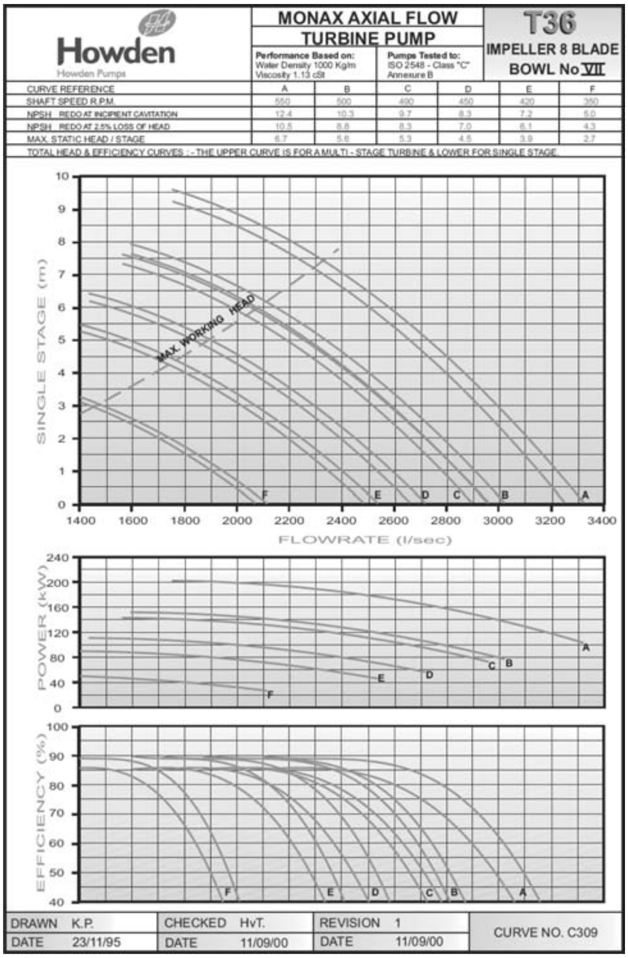


Fig. 4 Common pump curve.
Source: From Howden Pumps SA.

Positive Displacement Pump Theory

For positive displacement pumps, the power requirement of the pump is affected by flow rate, pressure, and fluid characteristics. The relationships used to determine performance for changes in total head, pump speed, or fluid density are given below in Eqs. 10–13.

$$Q_1 = Q_2\left(\frac{N_1}{N_2}\right) \tag{10}$$

$$W_1 = W_2\left(\frac{P_1}{P_2}\right) \tag{11}$$

$$W_1 = W_2\left(\frac{Q_1}{Q_2}\right) \tag{12}$$

$$W_1 = W_2\left(\frac{\delta_1}{\delta_2}\right) \tag{13}$$

where Q, flow rate (m^3/h); P, total head (m); W, power (kW); N, pump speed (RPM); and δ, fluid density (kg/m^3).

Pump Power Requirements

If one is to assume that a pump is 100% efficient, the theoretical power required for a specific flow rate against a given total head is determined by the relationship in Eq. 14.

$$W_t = \frac{QP_t}{367} \tag{14}$$

where Q, flow rate (m^3/h); P_t, total head (m); and W_t, theoretical power (kW).

The relation shown above can be expanded to incorporate the pump efficiency, belt drive efficiency, and motor efficiency. The incorporation of these variables results in Eq. 15.

$$W_{overall} = \frac{QP_t}{\eta_p\eta_d\eta_m 367} \tag{15}$$

Q, flow rate (m^3/h); P_t, total head (m); $W_{overall}$, theoretical power (kW); η_p, pump efficiency; η_d, belt drive efficiency; and η_m, motor efficiency.

The overall efficiency of the pump is given by the ratio of output energy to input energy in Eq. 16. This is the same the efficiency as given for the pump by the manufacturer.

$$\eta = \frac{\text{power output}}{\text{power input}} = \frac{\text{theoretical power}}{\text{actual power}} \quad (16)$$

Multiple Pumps Systems

Pumps can be installed in either parallel or series. For a series configuration, the combined pump curve is given by the sum of the heads of the individual pumps at equal flows. For pumps in parallel, the sum of the flows is taken at points of equal head. Typical curves for pumps in series and parallel are shown in Fig. 5.

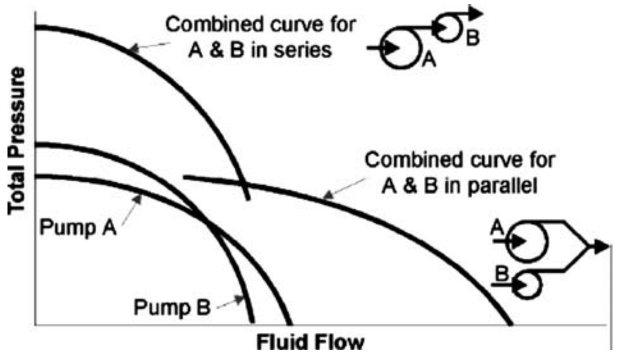


Fig. 5 Combined pump curves for series and parallel configurations.
Source: From ENEWS (see Ref. 1).

Determining of Power Usage

One of the most frequent requirements of energy conservation work on pumps is to estimate the change in the power required for a change of head or flow rate. In ideal situations, the performance curves of the pump are available. By superimposing the new system curve on the pump performance curve, the changes in head, flow rate, and power can be read directly from the pump curve.

In cases where the performance curves for pumps are not available, the calculations to determine the system curve can become very laborious. In such a situation, use can be made of the affinity laws. It must be kept in mind, though, that the answers resulting from the use of the affinity laws may not coincide with the actual system performance curve. The error, however, is small enough if the changes in pump settings are small.

Pump Seals

One of the factors that has a large effect on pump efficiency is the type of pump seal used and the quality of maintenance. The seal uses up a portion of the shaft power provided by the motor and the leakage caused by faulty seals represents a portion of the pumped fluid. The two most common types of pump seals are packing glands and mechanical seals. An example of each is in Fig. 6.

Packing Glands. Packing gland seals are made up of a set of rings, made of low friction material, compressed to achieve a close contact between the shaft and pump casing. The packing gland is cooled by forced lubrication. Common lubrication methods used are controlled leakage of pumped liquid, forced flushing by a separate liquid, and controlled force feeding of oil or grease. For any type of packing gland, the power used, the liquid lost, and life of the packing gland depends on the skilled adjustment of the pressure exerted on the packing rings by the retaining rings. A rule of thumb to keep in mind is that the power usage of packing glands is about six times greater than that of mechanical seals.

Mechanical Seals. Mechanical seals are made up of a set of rigid, spring-loaded rings, which slide against a finely finished mating surface. The rings are manufactured of low friction material, which is either self-lubricating or lubricated through a slight leakage of pumped liquid.

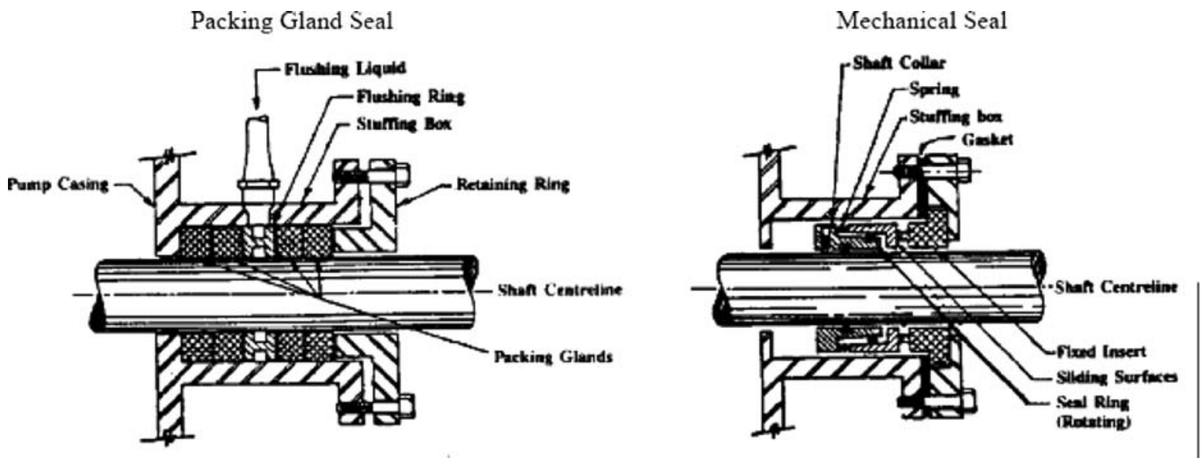


Fig. 6 Typical pump seals.
Source: From ENEWS (see Ref. 1).

Although the initial cost of mechanical seals is high, the lower power usage results in significant cost savings. For pumps that have packing glands, special types of mechanical seals are available as a retrofit option.

Energy Management Opportunities

Maintenance

Maintenance is one of the often-overlooked aspects of energy management. An effective maintenance program can result in viable savings within a very short time span. Some of the most important issues are given and discussed, cross-referencing the readings with the manufacturer's specifications.

Check Packing Glands. Packing glands should be checked regularly to ensure the glands are properly tightened and adjusted. For certain types of packing material, liquid temperature, shaft speed, optimal tightness, and allowable leakage may be established by monitoring the rate of leakage from the packing.

Check Critical Tolerances. Pump efficiency is affected by the leakage past the impeller from the intake (suction side) to the outlet (discharge side). On some pumps, use is made of replaceable wear rings, with small clearances between the moving surfaces, to minimize leakage and improve maintenance. The critical tolerances of mechanical seals stop leakage and air from being drawn into the fluid flow. The critical tolerances can be affected by the erosion of the impeller and wear rings if liquids are pumped that contain abrasive particles. So, to maintain high pump efficiency, the critical clearances and surfaces need to be checked periodically.

Check and Adjust Drives. If properly designed and maintained, drives such as belts or flexible couplings, can provide many years of hassle-free service. To ensure this, the following actions need to be carried regularly.

- Maintain alignment of pulleys and couplings
- Check tension of belts
- Lubricate bearings
- Replace or repair damaged belts, pulleys, clutches, drive keys, and couplings

The proper tension for various types of belts can usually be found in handbooks or catalogues available from component manufacturers.

Clean Pump Impeller. Pumps need to be cleaned regularly to maintain the efficiency of the pump. This is especially true of pumps that move "dirty" liquids. A build-up of dirt on the impeller and pump casing results in static pressure losses, which in turn reduces the efficiency of the pump.

Pump Scheduling. By switching off pumps that are not required to maintain the required fluid flow of the system, significant savings in energy and maintenance costs can be obtained. For example, hot water circulating pumps can be switched off in the summer, or process cooling water pumps can be switched of when the system in not working.

Establish Maintenance Program. Using the manufacturer's recommendations as guidelines, a maintenance plan should be set up. This maintenance plan should be designed for the use and type of pump. The program should typically include the following:

- Daily: Monitor pump sound, bearing temperature, leakage, and gauge readings.
- Semi-annually: Check and lubricate components in need of lubrication; check free movement of the stuffing box glands, and check the pump and drive alignment.
- Annually: Clean, inspect, and lubricate bearings and their seals, examine the packing and the shaft sleeve, recalibrate all associated instrumentation, and check performance against the design ratings.
- Replace worn components when test indicate a drop in performance.
- Adjust impeller clearance when test shows a loss of performance power requirement.

Retrofit Opportunities

In certain cases, opportunities arise which, if properly investigated and capitalized on through the use of a well-designed retrofit option, can result in viable savings of energy and costs. Some of the available retrofit options are discussed below.

Variable Speed Controller. Variable speed drives can be used to control the speed of the pump and also the flow rate and head. As such, the use of variable speed drives offers a potentially large saving in energy and cost for pumping systems.

If a throttle valve is removed from a pumping system, the system head curve will usually be lowered. This is due to elimination of the loss caused by the throttle valve, which can be as high as 25% of the total friction loss of the system. The reduction in the system head creates a potential for savings in energy if a new optimally sized pump is used. Although it is not always possible to replace the pump every time there is a change in the system curve, the installation of a VSD makes practical sense. A VSD-driven pump provides an infinite family of system curves depending on the speed.

In most cases, the high initial capital expenditure associated with a VSD can be justified; this is especially true of cases where the required system output varies over a wide range. If during the initial design of a new pumping system, the control valves and pump bypass piping is eliminated in favor of a VSD, the cost of the VSD will be defrayed. On existing systems, the cost of the retrofit will, in most cases, be paid back in less than two years. The

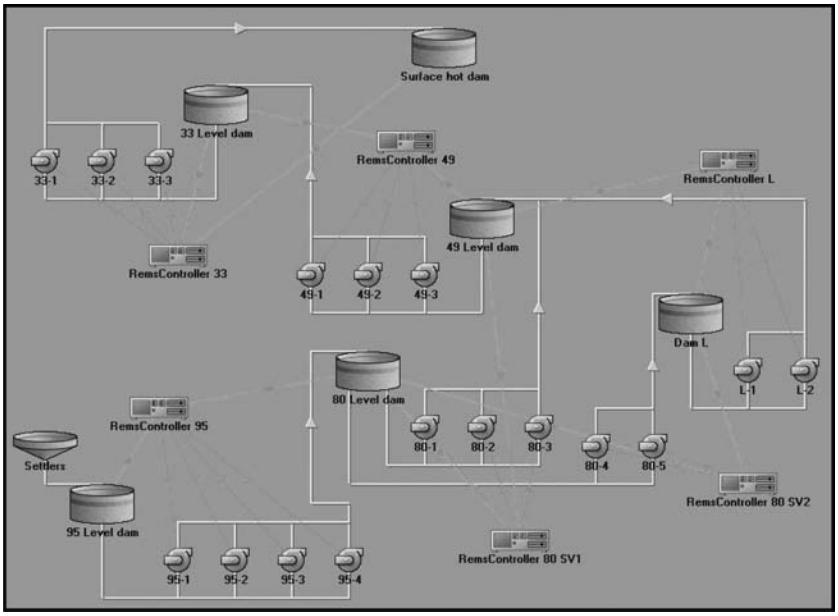


Fig. 7 Schematic layout of a typical intelligent pump scheduling system.
Source: From HVAC International.

savings and efficiency of the retrofit, however, is dependent on the type of VSD used.

Other advantages of the use of VSDs are:

- Improved process control
- Reduced wear on the pump and motor
- Reduced noise
- Reduced maintenance

For VSD applications, there is no single solution. Despite this, not all pump applications are good candidates for a VSD retrofit. Systems that typically fall in this category are:

- Applications with a low duty cycle
- Pumps that operate at a practically constant pressure
- Systems requiring a constant flow at a constant head

Intelligent Pump Scheduling. With this method, an automated scheduling system automatically schedules pumping activities out of peak periods. Extensive use is made of this method in South African mines on cold water drainage pump systems.

In this application, the system schedules the pumps to fill underground dams during the off-peak periods. This allows the pumps to be switched off during the peak period, resulting in cost savings. An example of such a control system is shown schematically in Fig. 7.

Although this method results in cost savings, the energy use remains the same. This is due to the fact that the energy use is just shifted from the peak to the off-peak periods.

Multiple Pumps. Savings can be had from using two smaller pumps in parallel instead of one large pump for systems where a wide range of flow rates is required at a reasonably constant pressure. For such a system, both pumps can operate simultaneously during peak demand periods, while only one would satisfy the need in lower demand periods. Energy savings result from avoiding the need to throttle the much larger pump during lower demand periods and from running the smaller pumps closer or on the optimum efficiency point.

Another option is to use one smaller pump in parallel to one or more larger pumps. This is a useful option where the capacity demand varies considerably. In such cases, the small pump can match the very lowest demand, while the larger pumps operate as normal.

The last option is to use one normally-driven pump in parallel with a VSD-driven pump. In such a setup, the VSD can be controlled until the desired capacity is matched. For very low demand, the normal pump can be switched off, while the VSD-driven pump is used to carry the load.

Power Recovery Turbines. In processes where high pressures are required, some of the residual energy can be used to drive a Hydraulic Power Recovery Turbine (HPRT) instead of dissipating it through a throttling valve. A good example is chilled water being fed down a mineshaft. The water is at a high pressure when it reaches the bottom, and instead of throttling the flow to avoid exceeding maximum pressure limits, an HPRT can be used to recover the latent energy contained in the water. This energy can be used to drive a generator, which in turn generates electricity to power either one of the pumps or some other application.

Another useful fact to remember is that practically all centrifugal pumps will perform as turbines when run in reverse.

Other methods that are not discussed above include:

- The replacement of outsized equipment
- Replacement of oversized motors

FAN SYSTEMS

A fan is defined by the Internet Free Dictionary as "a device for creating a current of air or a breeze, especially a machine using an electric motor to rotate thin, rigid vanes in order to move air, as for cooling."

Fans are one of the most misused, abused, and faultily applied types of equipment. The result of this misuse is high energy costs, which offers a large scope for the implementation of energy management methods, and accordingly also for sizeable savings.

Types of Fans

As with pumps, there are two basic types of fans—centrifugal and axial. These types are discussed in detail in the following paragraphs.

Centrifugal Fans

Centrifugal fans create a fluid flow through a centrifugal force, which is produced by moving a fluid between rotating impeller blades, and by the inertia generated by the velocity of the fluid leaving the impeller. Centrifugal fans are further differentiated by the shape of the fan blades. The four most common shapes are:

- Forward Curved
- Backward Inclined
- Airfoil
- Radial Blade

An example of each of the blade types, along with some typical performance curves, is shown in Fig. 8.

Fans with forward curved blades are used in low-pressure systems and run at a slower speed than backwardly inclined types. For this type of fan, a decrease in system static pressure at a constant operating speed results in a considerable increase in horsepower. This increase in horsepower could possibly lead to motor overload.

Backwardly inclined blades are used in high-pressure duct system applications. This type of blade is also non-overloading under normal conditions. This type of fan will have fewer blades on the wheel and these blades may be flat or slightly curved. This is in contrast with the forward curved fan on which the blades are more closely spaced and never flat. This type of fan also has a higher efficiency over the design range and operates at considerably higher speeds than the forward curved fan.

Airfoil fans are a special type of backward curved airfoil blade. They are used on all types of systems, from low to high pressure and run at very high speeds. They are very efficient, but this efficiency comes at a very high cost.

The last type is the radial blade. This type of fan is used for material handling and has self-cleaning blades. Radial bladed fans are the least effective type.

Axial Fans

Air is moved by an axial fan through the change in velocity of the air passing over the impeller. No energy is added to the fluid flow by the centrifugal forces. As with centrifugal fans, axial fans are distinguished by the shape of the fan blades. The most common types are:

- Propeller Fans
- Tubeaxial Fans
- Vaneaxial Fans

These types are shown in Fig. 9 along with common performance curves for some axial type fans.

The propeller fan usually is either belt or driven, and has either flat or slightly curved blades. This type of fan is used in applications associated with very low pressures and are seldom connected to a duct system. This type of fan is mostly used to move large volumes of air through large openings, which causes a small pressure loss.

Tubeaxial fans are capable of higher-pressure differentials and are more efficient than propeller fans. This type of fan typically has four to eight curved or airfoil type blades. The housings are made of cylindrical tubes, so that the clearance between the housing and blade tips is minimal. This type of construction results in the higher efficiency. The advantages of using tubeaxial fans are:

- Ease of installation
- Reasonable cost
- Low maintenance

Vaneaxial fans have blades with an airfoil design and can be operated against pressure. This type of fan is usually directly driven and has a higher efficiency than propeller type fans. For an increase in speed and/or system resistance, this type of fan will show a sharp increase in horsepower. This can result in motor overload.

Fan Measurements

Flow Measurement

The total pressure exerted by an air stream in a duct is the sum of the static pressure and the velocity pressure. A common device used to measure airflow is the Pitot tube shown in Fig. 10.

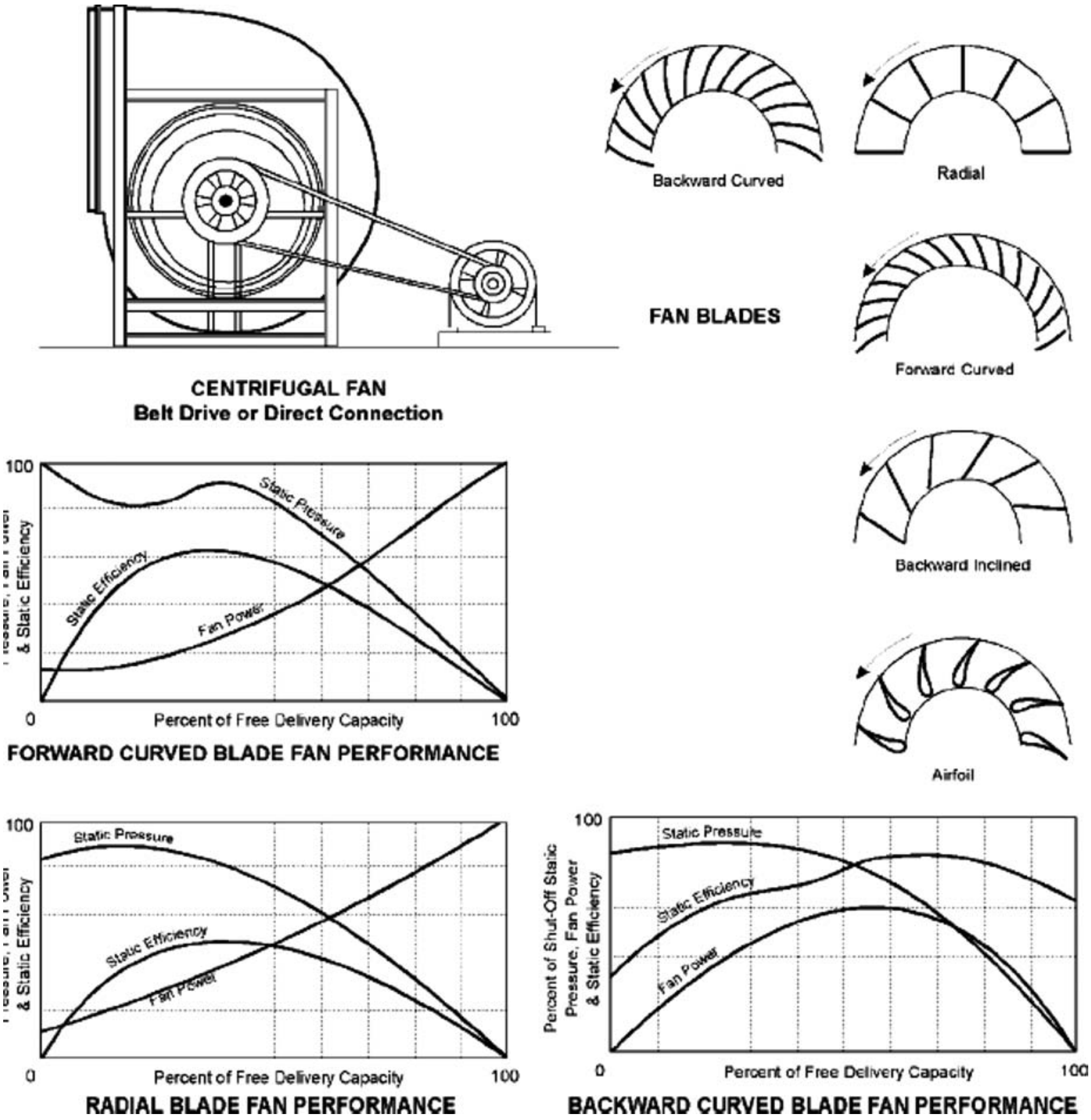


Fig. 8 Types of centrifugal fan blades.
Source: From ENEWS (see Ref. 1).

Measurements should be taken with the Pitot tube in accordance with the traverse detail shown in Fig. 10. Velocity pressure should be calculated for each of the traverse positions and then averaged. By using these measurements, the velocity of the flow in a duct can be calculated using Eq. 17.

$$\text{Velocity} = 0.764\left(\frac{TP_v}{B}\right)^{0.5} \tag{17}$$

where velocity, average air velocity (m/s); T, temperature (°K); P_v, velocity pressure (Pa); B, barometric pressure (kPa absolute); and 0.764, equation constant and unit conversion.

This equation can be simplified to take into account the standard conditions of 20°C and 101.325 kPa. This simplification results in Eq. 18.

$$\text{Velocity} = 1.3P_V^{0.5} \tag{18}$$

The volume airflow rate can now be calculated, using Eq. 19.

$$f_a = \text{Velocity } A_d 1000 \tag{19}$$

where f_a, volume air flow rate (l/s); A_d, area of duct (m^2); and 1000, conversion of units.

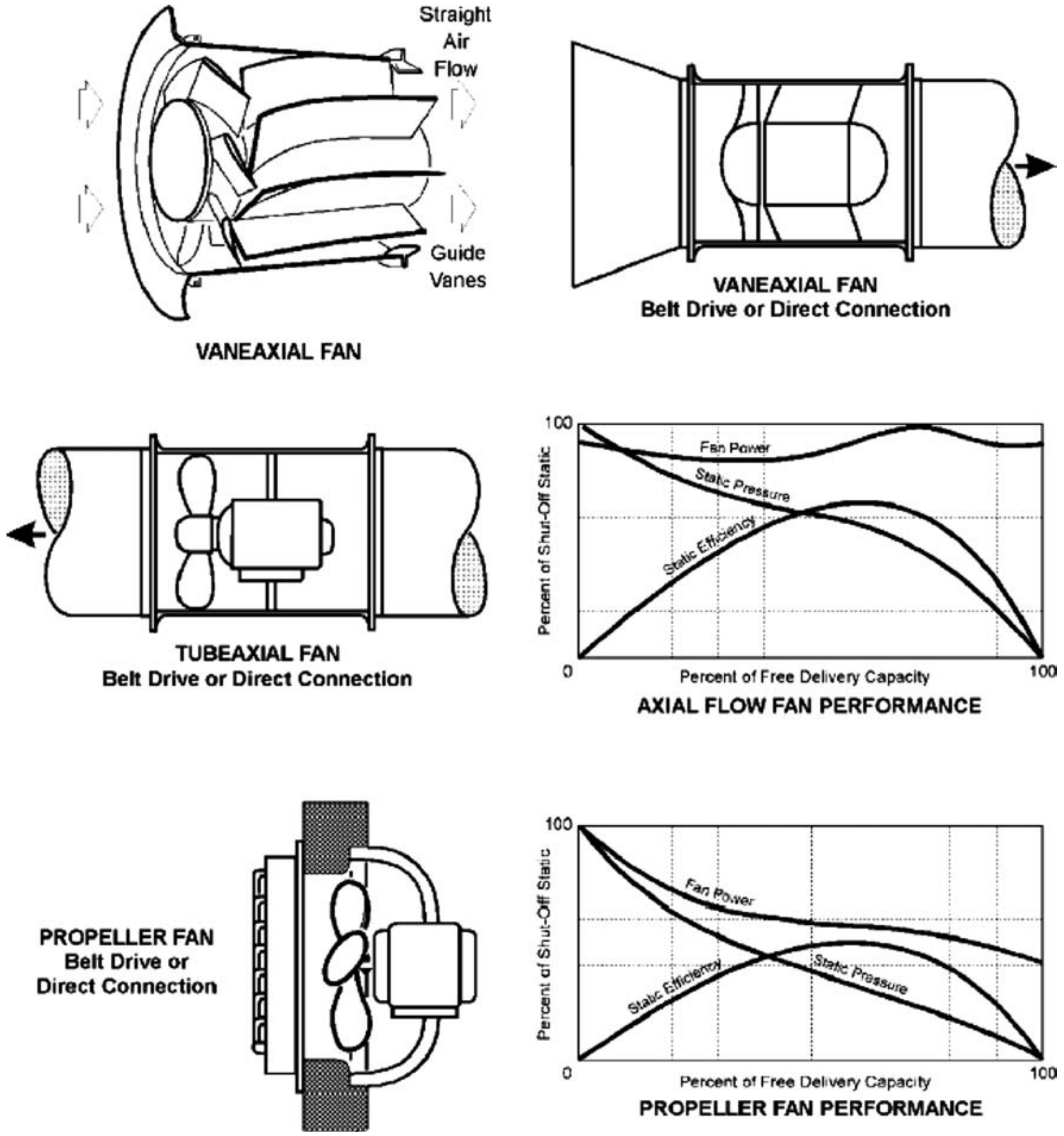


Fig. 9 Types of axial fan blades.
Source: From ENEWS (see Ref. 1).

Performance Measurement

Fan performance can be easily calculated using the measurements taken at the inlet and outlet through the method described in the previous section. There are several factors, however, which can affect the accuracy of these measurements. These factors are:

- Air flow not at right angles with the measurement plane
- Non-uniform velocity distribution
- Irregular cross sectional shape of the duct or passageway
- Air leaking between the measurement plane and the fan

For more precise measurements, refer to the Field Performance Measurements, Publication 203 by the Air movement and Control Association Inc. (AMCA).

Using the measured data, the total differential pressure across the fan can be calculated using Eq. 20.

$$DP_t = Ps_o + Pv_o - Ps_i - Pv_i \tag{20}$$

where DP_t, total differential pressure (Pa); Ps_o, static pressure at outlet (Pa); Pv_o, velocity pressure at outlet (Pa); Ps_i, static pressure at inlet (Pa); and Pv_i, static pressure at inlet (Pa).

The total fan static differential pressure can also be calculated using Eq. 21

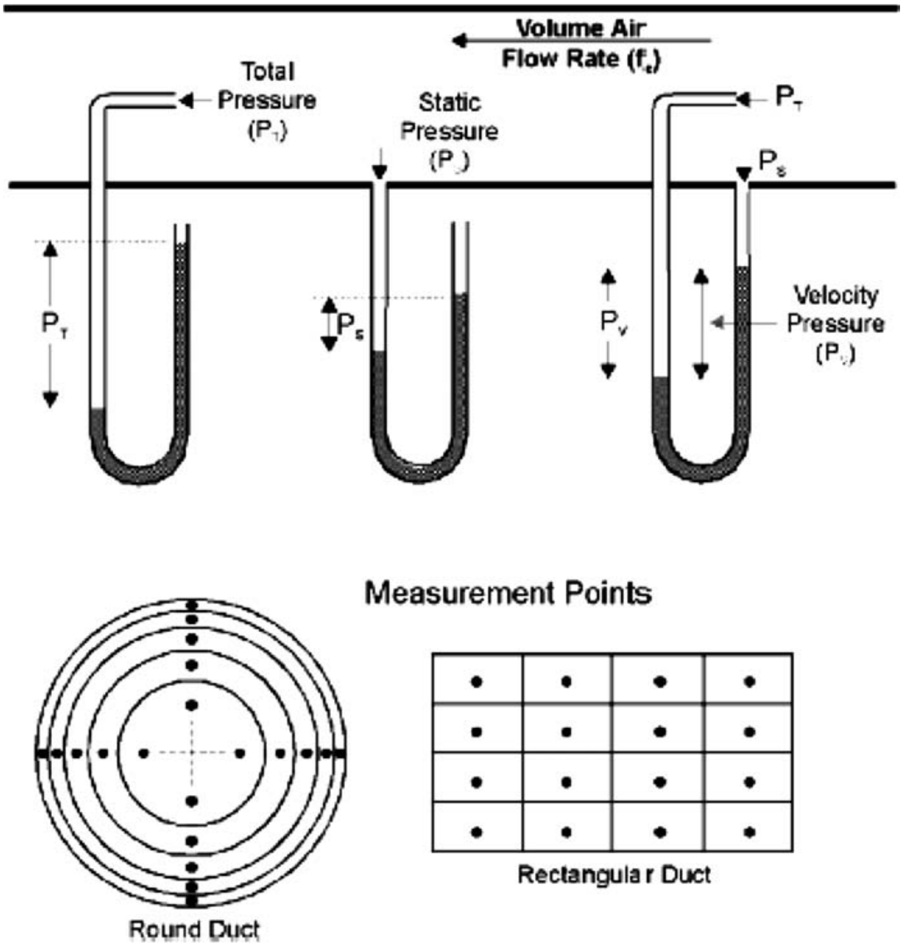


Fig. 10 Pitot tube measurements. Source: From ENEWS (see Ref. 1).

$$DP_s = Ps_o - Ps_i - Pv_i \quad (21)$$

where DP_s, total fan static differential pressure (Pa).

Although the effects of inlet and outlet conditions have not been included in the equations, the equations still provide a reasonable basis for the further calculation of static efficiency and power.

Fan Performance

Power Requirements

By measuring the quantity of air delivered, and the pressure against which it is delivered, it is possible to calculate the work done by a fan. If one is to assume that the fan is 100% efficient, the required theoretical power needed to move a volume of air against a total pressure can be calculated using Eq. 22.

$$W_t = \frac{QP_t}{1000} \quad (22)$$

where W_t, theoretical power to move air at 100% efficiency (kW); Q, static pressure at inlet (Pa); and P_t, static pressure at inlet (Pa).

The overall fan efficiency relates the theoretical output power of the fan against the input electrical power. This relation includes fan losses, belt drive losses, and motor losses.

Performance Curves

As with pumps, information on the performance of commercial fans is available from the manufacturers in the form of performance curves. An example of a performance curve is shown in Fig. 11.

A fan operates along a specific performance curve for a specific speed. Thus, at a fan speed of N_1, the fan will operate along the N_1 performance curve as shown in Fig. 11. The actual operating point on the curve is dependent upon the system's operating characteristics.

For any fan system, pressure increases with an increase in airflow. The pressure requirement of a system over a range of flows can be determined, and using this data a "system performance curve" can subsequently be developed (shown as SC_1 in Fig. 11). If the system curve is plotted on the fan curve, the point where the two curves intersect is the actual operating point of the fan (shown as point "A" on Fig. 11).

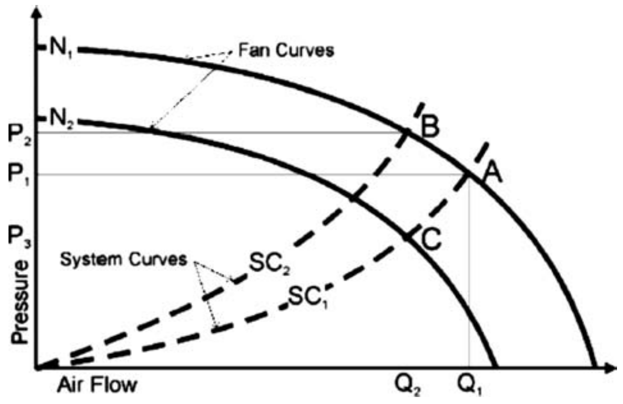


Fig. 11 Fan performance curve.
Source: From ENEWS (see Ref. 1).

Two methods can be used to reduce airflow. One method is to restrict the flow by partially closing a damper. This causes a new system performance curve (SC_2 on Fig. 11) to develop where the required pressure is greater for the new given airflow.

The second method is to reduce the speed of the fan (N_2 on Fig. 11) while keeping the damper fully open. The fan would now operate at point "C" to provide the same airflow, but at a lower pressure. The reduction of fan speed is a more efficient way to decrease airflow, since it requires less power and thus less energy is used.

Impact of Air Density

The performance data provided by manufacturers is based, unless otherwise specified, on dry air at a standard atmospheric pressure (101.325 kPa), temperature (20°C), and density (1.2 kg/m^3).

In most applications, moist air at temperatures and pressures other than the standard conditions are processed. Because of this, the air density needs to be corrected to obtain the actual fan performance. By measuring the wet bulb, relative humidity, or dew point temperature, the moisture content of an air stream can be determined. Wet bulb or dry bulb temperatures are mostly determined at the fan inlet using a sling thermometer. When the air stream temperature is above 82°C, a more reliable determination of moisture content is made from the dew-point temperature. If the dry bulb temperature is between 5 and 38°C, density may be determined by using a psychrometric chart.

Fan Laws

Fan laws are very similar to the laws that are used to predict pump performance. The laws are made up of a set of relations concerning speed, power, and pressure. A change in RPM on any fan causes a predictable change in pressure and power. The basic fan laws are shown in Eqs. 23–25.

$$\frac{Q_1}{Q_2} = \frac{N_1}{N_2} \tag{23}$$

$$\frac{P_1}{P_2} = \left(\frac{N_1}{N_2}\right)^2 \tag{24}$$

$$\frac{kW_1}{kW_2} = \left(\frac{N_1}{N_2}\right)^3 \tag{25}$$

where W, power (W); N, speed (RPM); P, pressure (Pa).

Energy Management

Air handling systems are usually balanced after installation, which requires the services of an experienced specialist. For this article, it is assumed that the air handling system under consideration has been balanced after installation. However, during the operating lifetime, the components comprising the air handling system wear down.

Dirt accumulated on fan blades can cause the blades to become unbalanced, which in turn causes vibrations and noise. Faulty bearings can also cause vibrations and damage to the fan blades. Fouled filters reduce airflow, and thus reduce motor load, and can create a false sense of savings. From this, it is apparent that the first step, as with pumps, is to bring air handing systems up to standard and to implement a proper maintenance schedule.

Maintenance

Some common maintenance actions are listed and discussed in the following paragraphs.

Check and adjust belt drives regularly. Misaligned drive sheaves damage belts and can increase the power requirement of the fan. In Fig. 12, drive losses for properly aligned drives are shown for different speed ranges. Losses due to misaligned drives fall outside the limits set in Fig. 12.

Lubricate fan components according to the manufacturer's instructions. Fan components, such as couplings, shaft bearings, adjustment linkage, and adjustable supports, need lubrication with a proper lubricant at intervals recommended by the manufacturer.

Clean fan components regularly. All fans should be cleaned regularly to maintain efficiency. This is especially true of fans handling dirty air. Dirt that has accumulated on the fan blades and housing causes higher static pressure losses, and these losses in turn reduce efficiency.

Correct excess noise and vibration. Fan noise and vibration can be caused by a number of factors:

- Out of balance fan wheel
- Bad bearings
- Insufficient insulation

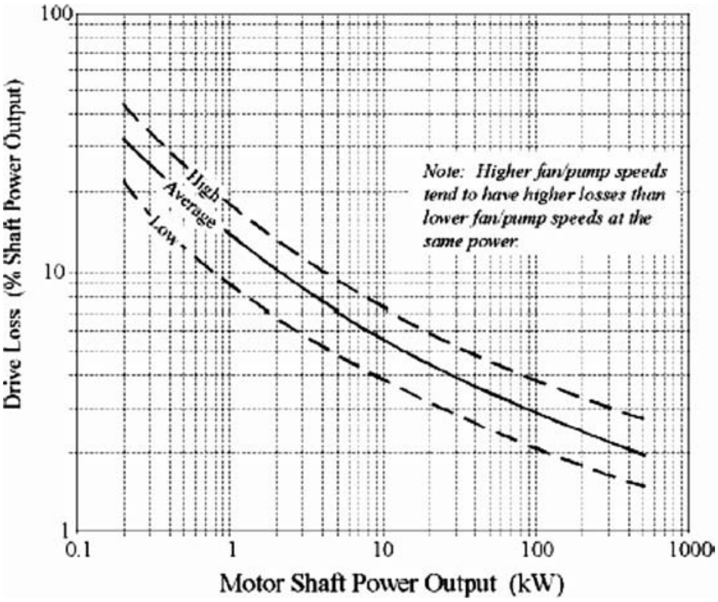


Fig. 12 Drive losses.
Source: From ENEWS (see Ref. 1).

- Misalignment of shaft seals
- Corrosion between shaft and bearing

Correct leaks. Energy savings can be attained by fixing air leaks from loose connections, improperly sized dampers and shaft openings, and unsealed expansion connections.

Replace loaded air filters. One common cause of poor system performance is loaded air filters. Manufacturers provide data on the point at which a filter is considered fully loaded. This data is rated in terms of pressure drop at various velocities. When it is found that the pressure drop over a filter reaches a point where the filter becomes loaded, the filter should be replaced. Furthermore, if a system with loaded filters is balanced, and the filters subsequently replaced, the airflow might become excessive, which would result in higher system energy.

Implement a maintenance program. Implement a maintenance program similar to that described for pumping systems above.

Low Cost Options

These are energy management options that are relatively low cost and can be implemented easily. Some low cost options worthy of evaluation are:

- Improve fan inlet and outlet duct connections to reduce losses at the inlet and outlet
- Shut off fans that are not needed
- Reduce fan speed to provide the optimum amount of airflow when the dampers are open in the maximum position for balanced air distribution

Retrofit Options

The implementation cost of any retrofit option is significantly higher than low cost options. Below are some typical energy management options that are available under the Retrofit option:

- Use a variable speed motor to allow the fan to adjust for changes in system requirements.
- Replace oversized motors.
- Replace outdated equipment with newer, optimal sized units.
- Install digital control system.

Flow Control

Most air handling systems are designed for 100% airflow. However, in some cases, they can operate at lower flow rates. What options are available to a manager to achieve a lower flow rate? The following options are available.

- Use dampers or inlet vanes to control the flow.
- Install a variable speed drive.

Fig. 13 shows the typical power required at a specific flow rate for different control methods. As most fans are usually oversized, the curves are based on the power requirement of a fan which is 10% oversized. Most fans are oversized because of:

- Leaks that develop and filters clogging
- Conservative design engineers
- The fan's inability to meet design specifications

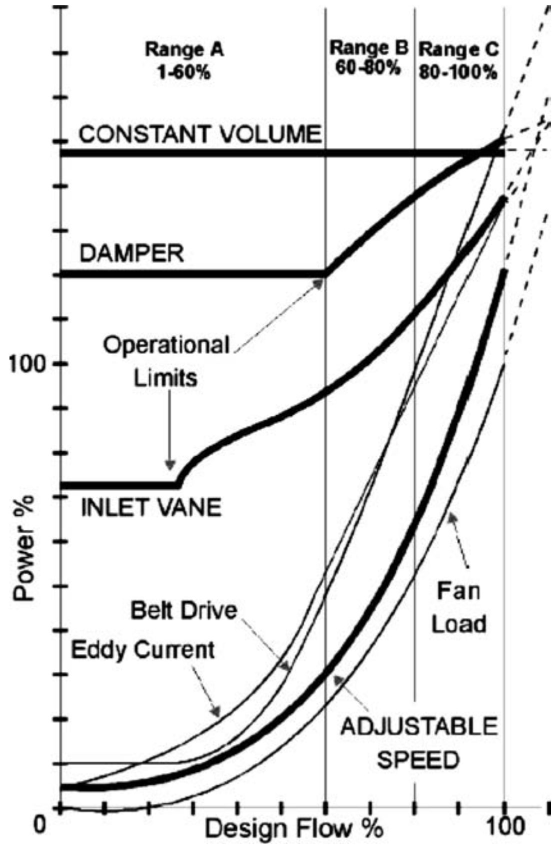


Fig. 13 Typical power to flow curves for different control methods.
Source: From ENEWS (see Ref. 1).

Furthermore, the curves are also based on AC motors running at full load and 90% efficiency, and with a typical efficiency drop for reduced loads.

There are, however, some limitations associated with each of these methods. Dampers have a limited turn down and may be become noisy. Inlet vanes, even if wide open, restrict flow. Although variable speed drives reduce noise and save energy by adjusting the speed of the fan to the exact needs of the system, they are very expensive, and the potential savings may not justify the initial capital expenditure.

CONCLUSION

This article focused on energy management opportunities that exist within pumping and fan systems. It provided an overview of the types of pumps and fans, their operating characteristics, theory, and power requirements, and provided typical cost and energy saving opportunities.

REFERENCE

1. SADC Industrial Energy Management Project. Core Training Manual. Zimbabwe.

BIBLIOGRAPHY

1. Reeves, G.; Gillespie, K.L.; Cowan, J.D.; et al. *ASHRAE Guideline 14-2002—Measurement of Energy and Demand Savings*; American Society of Heating, Refrigeration and Air-Conditioning Engineers: 1791 Tullie Circle, NE - Atlanta, GA, 2002.
2. Cooper, P. Pump Machinery. In Cooper, P., Ed.; American Society of Mechanical Engineers: United Engineering Centre, 345 East 47 Street, New York, 1989.
3. Warring, R.H. *Pumping Manual 7th Ed.*; Gulf Publishing Company: Houston, TX, 1984.
4. Karnassik, I.J.; Messina, J.D.; Cooper, P.; et al. *Pump Handbook; 3rd Ed.*; McGraw Hill Book Company: Singapore, 2001.
5. Hydraulic Institute. *Energy Reduction in Pumps and Pumping Systems*; Hydraulic Institute: U.S.A., 1997.

Radiant Barriers☆

Ingrid Melody
Philip Fairey
David Beal
Florida Solar Energy Center, Cocoa, Florida, U.S.A.

Abstract
Attic radiant barriers made of aluminum foil are becoming a popular way for homeowners to save energy and money in Southern states. They are increasing in popularity for two reasons. First, tests by the Florida Solar Energy Center (FSEC) and other groups show that they work. Second, manufacturers are improving the quality of radiant barrier materials.

To most homeowners, attic radiant barriers are a new energy conservation concept; many of them have questions about how radiant barriers work and how to use them. Herein are answers to some of the most commonly asked questions. Also included are recommended ways to install radiant barriers in existing attics and new homes.

RADIANT BARRIERS

A radiant barrier is a layer of aluminum foil placed in an airspace to block radiant heat transfer between a heat-radiating surface (such as a hot roof) and a heat-absorbing surface (such as conventional attic insulation). Fig. 1 illustrates the locations in which a radiant barrier may be installed in an attic. Only locations 1 and 2 are recommended for sheet radiant barriers because dust will accumulate on the radiant barrier if it is installed at location 3.

BENEFITS

In hot climates, benefits of attic radiant barriers include both dollar savings and increased comfort.

Without a radiant barrier, your roof radiates solar-generated heat to the insulation below it. The insulation absorbs the heat and gradually transfers it to the material it touches, principally, the ceiling. This heat transfer makes your air conditioner run longer and consume more electricity.

An aluminum foil radiant barrier blocks 95% of the heat radiated down by the roof so it can't reach the insulation.

In summer, when your roof gets very hot, a radiant barrier cuts air-conditioning costs by blocking a sizable portion of the downward heat gain into the building.

In the warm spring and fall, radiant barriers may save even more energy and cooling dollars by increasing your personal comfort. During these milder seasons, outdoor air temperatures are comfortable much of the time. Yet solar energy still heats up your roof, insulation, attic air, and ceiling to temperatures that can make you uncomfortably warm. An attic radiant barrier stops almost all of this downward heat transfer so that you can stay comfortable without air conditioning during mild weather.

You may also find that radiant barriers can expand the use of space in your home. For instance, uninsulated, unconditioned spaces such as garages, porches, and workrooms can be more comfortable with radiant barriers. In addition, because radiant barriers keep attics cooler, the space is more usable for storage.

One final benefit: a cooler attic transfers less heat into air conditioner ducts, so the cooling system operates more efficiently.

"BLOCKING" HEAT TRANSFER

Aluminum foil, the operative material in attic radiant barriers, has two physical properties of interest here. First, it reflects thermal radiation very well. Second, it emits (gives off) very little heat. In other words, aluminum is a good heat reflector and a bad heat radiator.

Your grandmother probably made use of these properties through "kitchen physics." She covered the Thanksgiving turkey with a loose "tent" of aluminum foil before she put it in the oven. The foil reflected the oven's thermal radiation, so the meat cooked as evenly on top as on the bottom. She removed the foil briefly to let the skin brown, but when she took the bird from the oven, she "tented" it with foil again. Since aluminum doesn't emit

Keywords: Radiant barriers; Hot roofs; Attic insulation; Infrared heat; Low emissivity; Radiant heat transfer; Aluminum foil barriers.

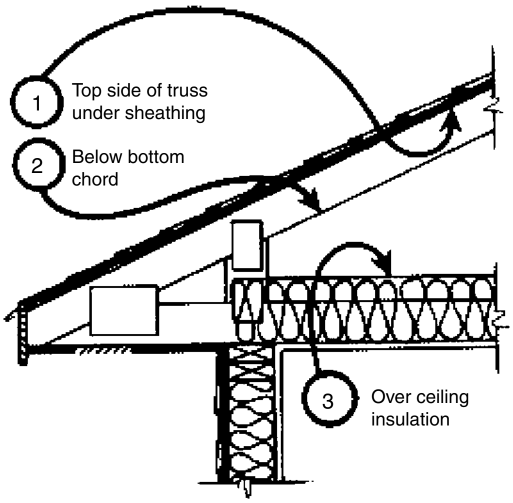


Fig. 1 Typical attic section with three possible locations for radiant barrier.

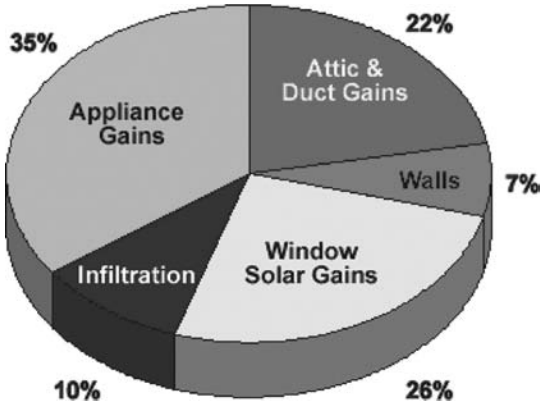


Fig. 2 Breakdown of cooling loads in a typical 1500-ft^2 central Florida home.

much heat, the turkey stayed hot until the rest of the meal was ready. To understand the concept of not emitting heat, let's use an analogy of a light bulb. When you turn on a light bulb, it emits light. If you were to paint the light bulb black, when it was turned on, it would not emit light. A radiant barrier has a similar effect on infrared heat. Your roof surface heats up in the sun and will emit infrared heat. When this infrared heat heats the radiant barrier it will not emit, or reradiate, the heat into your attic.

Cooking a turkey and painting light bulbs are simple analogies, but the same principles of physics apply to an attic radiant barrier. Aluminum foil across the attic airspace reflects heat radiated by the roof. Even if the radiant barrier material has only one aluminum foil side and that side faces down, it still stops downward heat transfer because the foil's low emissivity will not allow it to radiate the roof's heat to the insulation below it.

SAVINGS

Since everyone's home and lifestyle are different, we can't precisely calculate your personal savings from attic radiant barriers. However, it's reasonable to expect that an attic radiant barrier can save 8–12% of your annual cooling costs in the Southeast.

Savings from an attic radiant barrier depend on the amount of heat the roof and attic contribute to your home's cooling load. ("Cooling load" is the total amount of heat your air conditioner must remove to maintain comfortable indoor temperatures.) In general, the more energy efficient the rest of your home is, the larger the percentage of energy you save from an attic radiant barrier because the roof and attic make up a larger portion of the cooling load.

Fig. 2 shows a breakdown of cooling loads in a typical 1500-ft^2 Central Florida home. The attic (including heat gains to the duct system) accounts for 22% of the total cooling load. In this house, an attic radiant barrier could save 8–12% on the annual air-conditioning costs.

Although not as significant, heating savings may also accrue from the use of radiant barriers.

Results from a recent comprehensive field monitoring study conducted for Florida Power Corporation (FPC) by FSEC on the performance of attic radiant barrier systems in Central Florida homes may be viewed at http://www.fsec.ucf.edu/bldg/pubs/rbs/.

Claims of Greater Savings

As in most cases, claims for radiant barriers that *sound* too good to be true *are* too good to be true. If your roof accounts for less than 20% of your cooling load, then an attic radiant barrier can't possibly save more than 20% on your bills.

Claims of greater savings may simply be the results of misunderstanding. For instance, FSEC has measured and reported that radiant barriers can reduce heat gain through R-19 insulated ceilings by over 40%. If the ceiling portion of the total cooling load is 20%, that's a reduction of 40% of 20%, which amounts to 8% savings on the total cooling load.

If all the facts and figures tend to confuse you, just remember that an attic radiant barrier can save about 8–12% on your air-conditioning costs in the Southeast. Any Sunbelt homeowner knows that an 8–12% saving on air-conditioning bills can be significant.

AVAILABLE MATERIALS

There are many types of radiant barrier materials on the market, and more are being developed as radiant barriers become more widely used. Five generic types are most common:

- Single-sided foil (one foil side) with another material backing such as craft paper or polypropylene. Some products are further strengthened by fiber webbing sandwiched between foil and backing. The strength of the backing material is important since unreinforced foil tears very easily.
- Foil-faced roof sheathing materials that come from the manufacturer with a foil facing adhered to one side of the sheathing.
- Double-sided foil with reinforcement between the foil layers. Reinforcement may be cardboard, craft paper, mylar or fiber webbing.
- Foil-faced insulation. The insulating material may be polyisocyanurate, polyethylene "air-bubble" packing or other materials that impede heat conduction.
- Multilayered foil systems. When fully extended and installed so that the foil layers do not touch, these products also form insulating airspaces.
- Radiant barrier "chips" are also manufactured and sold. This product is slightly different than a conventional sheet-type radiant barrier in that the "chips," which are blown onto the floor of the attic—typically to a depth of 3 or more inches, act as a multilayer product with many "trapped" air pockets. These air pockets cause this product to function somewhat like traditional, fibrous insulation products. Even though this product may collect dust on its uppermost layer, the remaining layers and air spaces work to significantly reduce heat transfer through the ceiling assembly.

Some of these products may have R-values, which may be claimed only if the product was tested according to Federal Trade Commission regulations for insulation.

Although it is not by definition a radiant barrier, there is a low-emissivity paint available that can be applied directly to the underside of the roof decking.

Best Materials

While the Florida Solar Energy Center strongly recommends radiant barrier systems in attics, it doesn't endorse any particular brand of radiant barrier material. However, we suggest that you look for a few common-sense characteristics:

- Emissivity (the lower the better).
- Fire rating (as required by building codes).
- Ease of handling.
- Strength of reinforcement.
- Width appropriate for installation.
- Low cost.

COSTS

Costs for an attic radiant barrier depend on several factors, including the following:

- Whether purchase includes installation (which increases cost).
- Amount purchased (greater quantities can cost less per square foot).
- Manufacturing method and type of reinforcement.
- Presence of other insulation materials.
- Marketing methods.
- Aspects of supply and demand.

One other condition greatly affects the cost of a radiant barrier system to the homeowner—the individual's knowledge and willingness to do some comparison shopping. A few phone calls and a little research can save you money on most purchases. Radiant barriers are no exception.

Informal surveys show a wide range of material costs ($0.07–$1.00/ft^2) and installation costs ($0.10–$1.00/ft^2). The increases costs appear to be due more to marketing practices than to any inherent difference in thermal performance.

In some cases, radiant barriers are included in a package of energy-saving features sold to homeowners. When considering a "package deal," you may want to ask for an itemized list that includes material and installation costs for all measures included. Then shop around to see what each item would cost if you purchased them individually. You may see considerable savings. In addition, you may decide that you want to install the items yourself, including the radiant barrier.

FOIL SIDE

In attics, single-sided radiant barrier material should be installed with the foil side facing down. This may run counter to our intuitive feel for "how things work," but it does work, and work well.

To understand how it works, remember the two properties of aluminum foil from our Thanksgiving turkey and light bulb analogies; foil reflects radiant energy very well but does not radiate heat well. It does not emit heat to the cooler surfaces around it.

If you install a single-sided radiant barrier with the foil side facing up, the aluminum will (for a time) reflect the thermal energy radiated by the hot roof.

If you install a single-sided radiant barrier with the foil side facing down, the aluminum simply will not radiate the heat it gains from the roof to the cooler insulation it faces.

At first, a single-sided radiant barrier will work equally well with the foil facing up or down. However, over time, dust may accumulate on the surface of foil facing up. The dust will reduce the radiant barrier effect by allowing the foil to absorb rather than reflect thermal radiation. However, a radiant barrier with the foil side facing down will not collect dust on the foil and will continue to stop radiant heat transfer from the hot roof to the insulation over the life of the installation.

Even if you use a double-sided radiant barrier material, it is best to install it at the rafter level so that the bottom side faces the attic airspace and will not collect dust.

INSTALLATION

The most effective way to install a radiant barrier in an existing attic is simply to staple the foil material to the underside of the top chord of the roof trusses or to the underside of the roof decking.

It is not very easy to work in any attic, even one with a steep pitch. Always keep in mind that a misstep could be disastrous, since most attic "floors" are not floors at all, but rather 2×4s holding ceiling drywall topped by conventional insulation. You should consider safety first (see installation safety tips listed below).

Take care to avoid compressing existing insulation in the attic.

Tools and materials needed to install a radiant barrier include the following:

- Enough radiant barrier material to cover the underside of the roof.
- Measuring tape and flashlight.
- Heavy-duty scissors or utility knife.
- Staple guns and heavy-duty staples.
- Two movable support surfaces such as 3×2-ft sheets of 1 in. plywood or 3 ft lengths of 1×12 board.

Perhaps your most important aid will be a partner. Working in pairs in the attic makes the work go faster. Even more important, it adds to safety.

Begin by measuring the length of the attic roof from peak to soffit. Then, return to a stable, ground-or floor-level surface to measure and cut the radiant barrier material to size. The material usually comes in rolls of 50 to several hundred feet; it's easiest to cut and reroll all the lengths you'll need before returning to the attic.

At one end of the attic, place the plywood or 1×12 as a stable surface across two of the attic truss members. Try to minimize compression of existing insulation. Provide one surface at the peak and one at the soffit end so that both installers can work together.

Safety reminder: Be extremely careful at the sides of your support surface. If you step on an edge, the surface can tilt and drop you through the ceiling drywall below.

With your partner, unroll one length of the radiant barrier material from soffit to peak. Leaving 1 or 2 in. of free space at the roof peak, staple one corner of the material to the underside of the top chord of the first roof truss. Continue stapling the edge of the radiant barrier material down the truss at 6–12-in. intervals, stopping 2–3 in. from the ceiling insulation. Next, staple the other edge of the material to the underside of the adjacent roof truss. Continue the process at adjoining trusses until the underside of the roof is no longer visible except for a 1- or 2-in. strip at the roof peak.

As an alternative, you may staple the radiant barrier material to the underside of the roof decking, adjacent to the top chord of the truss. The weight of the material will allow it to drape naturally between trusses.

Safety Tips for Installation

- If you use a ladder for access to the attic, make sure it is stable and tall enough for easy entry and exit.
- Work in the attic only when temperatures are reasonable. Attic daytime temperatures can rise far above 100°F during much of the year in the Sunbelt. Install your radiant barrier system early in the morning, or wait until cool weather sets in.
- Work with a partner. Not only does it make the job go faster, it also means that you'll have aid should a problem occur.
- Watch where you walk and use a movable support surface. Step only on the attic trusses or rafters and your working surface. Never step on the attic insulation or the ceiling drywall below it.
- Step and stand only on the center of your movable working surface. Don't step on the edge; it can cause the surface to tip.
- Watch your head. In most attics, roofing nails penetrate through the underside of the roof. If you bump your head, it can cause a serious cut or puncture. If your skin is punctured by a nail, an up-to-date tetanus vaccination is a must. Avoid potential problems by wearing a hard hat.
- Be especially careful around electrical wiring, particularly around junction boxes and older wiring. Never staple through or over electrical wiring.
- Make sure that the attic space is well ventilated and well lighted. Bring in fans and extra work lights if necessary.
- If your attic has blown-in insulation, direct fans upward, away from the insulation material.
- Avoid exposure to mineral fiber insulation. Wear goggles, long pants, a long-sleeved shirt, and a particle mask or kerchief over your nose and mouth. Wear gloves if you are particularly sensitive to fiberglass.
- Wear a tool belt or utility apron to carry staples, staple gun, scissors, measuring tape, etc.
- Take frequent breaks, and pace yourself. It's better to get the job done over a longer period than to risk an accident due to fatigue or to end up with a poor-quality installation.

Airtightness

You're installing a barrier against radiated not convected heat, so you need not cut off air motion. In fact, ventilation from soffit to peak improves radiant barrier system performance.

Small tears and holes will not significantly lessen the performance of the radiant barrier, so don't worry if you must cut and patch around obstructions such as vent stacks and truss supports.

Placement

It's not recommended to place the material directly on top of insulation. In this type of installation, dust will accumulate on the foil surface facing the roof. In time, the dust will negate the radiant barrier effect. In addition, problems could develop with moisture condensation. An exception to this is the "chip" and multi layer products on the market. Their multiple layers give them resistance to the effects of duct build up, making them suitable for floor installations.

Installation in New Construction

There are two widely accepted methods of installing attic radiant barrier systems in new construction:

- Attach the radiant barrier to the roof decking before it is installed on the trusses, or
- Attach the radiant barrier to the roof deck or truss chords after the roof decking is installed, but prior to the installation of the ceiling drywall.
- Perhaps the easiest way to install a radiant barrier in a new house is to use one of the foil-faced roof sheathing materials on the market. The price increase for these products should be minimal, and the installation does not differ from regular decking, except for a little additional care in the handling of the product.

Note: When installing a single-sided radiant barrier material, remember to face the foil side down toward the attic floor/insulation.

FOIL-FACED BATT INSULATION

While some conventional batt-type insulations have an aluminum foil backing, it's probably not a good idea to simply flip your insulation over to use it as a radiant barrier. Not only will you encounter the eventual dust problem, you may also encounter a fire hazard in the glue that bonds the foil to the batt.

At least one batt insulation manufacturer has introduced a product with a foil face that is bonded to the insulation with a fire-retardant glue. This product meets fire codes, but it still has the potential for dust problems.

HEAT BUILDUP

The Florida Solar Energy Center has measured the temperatures of roof shingles above attic radiant barriers on hot, sunny summer days. Depending on the color of the shingles, their peak temperatures are only 2–5°F higher than the temperature of shingles under the same conditions without a radiant barrier.

Roofing materials are manufactured to withstand the high temperatures to which they are frequently exposed. A 2–5°F increase in peak temperatures that normally reach 160–190°F should have no adverse affect.

SHINGLE WARRANTIES

Shingle warranties should not be subject to cancellation by the manufacturer on the basis of radiant barrier installation. However, it may be wise to review the warranty to be sure that work of this nature will not void it. You may want to inquire directly of the manufacturer. Any changes in warranty should be substantiated in writing.

Reshingling

Remember, to be a radiant barrier, the aluminum foil must be installed facing *an airspace*. If there is no airspace, the foil acts as a conductor and quickly passes heat by conduction from a hot surface to a cooler one.

If your re-roofing job requires decking replacement, however, it is a good time to consider a radiant barrier, especially the foil-faced roof sheathing materials on the market.

DECREASING HEAT GAIN

While a radiant barrier is one effective way of reducing heat gain through attics, it's not the only one. Other options include:

- Continuous peak and soffit or gable vents (which also improve radiant barrier system performance).
- Light-colored shingles or white or light metal or tile roof coverings.
- Additional conventional insulation.

If you shop carefully, you will probably find that attic radiant barriers are one of the least costly and yet most effective of the attic conservation measures for Southern climates.

PAYBACK

Two things affect the performance of a radiant barrier system—the level of insulation in the attic, and the geographic location of the home. A radiant barrier system reduces the heat flow into the house from the attic by approximately 40%. Attic insulation levels have a large effect on the amount of heat flow that is reduced, in other words, if you have little or no insulation in your attic, a 40% reduction is very significant, but if your attic is insulated to R-30 or better, there is very little heat flow to reduce. The more of your energy bill that is concerned with heating, the less desirable having a radiant barrier becomes. When you are using your heater, any heat gain from the attic is desirable. There may be a reduction of heat loss through the roof during winter nights, but the climates where testing has been performed do not lend themselves to demonstrating this, as there is very often a brief or non-existent winter. A very helpful website for guidance to the cost effectiveness of installing a radiant barrier is found on Oak Ridge National Laboratory's web site, http://www.ornl.gov/sci/roofs+walls/radiant/rb_02.html.

In Florida, computer studies conducted in the development of the Florida Model Energy Code indicate that a typical attic radiant barrier installed in a Florida home will offer a 6–7 yr simple payback and a 15–19% return on investment.

CONCLUSIONS

Attic radiant barriers are an inexpensive but effective way for Sunbelt homeowners to save energy and money. While they are not a new concept, radiant barriers have only recently been proved effective for energy conservation.

Manufacturers are continuing to improve radiant barrier materials, which are becoming widely available throughout the southern states.

A radiant barrier may be installed in an existing attic or during construction of a new home. Both are relatively easy procedures.

ACKNOWLEDGMENT

Publication of this document was supported in part by Florida Power & Light Company.

BIBLIOGRAPHY

1. Parker, D.S.; Sonne, J.K.; Sherwin, J.R. Flexible Roofing Facility: 2002 Summer Test Results, Prepared for: U.S. Department of Energy Building Technologies Program, July 2003.
2. Parker, D.S.; Sherwin, J.R.; Anello, M.T. *FPC Residential Monitoring Project*: New Technology Development—Radiant Barrier Pilot Project, Contract Report FSEC-CR-1231-01, Florida Solar Energy Center, Cocoa, Florida, January 2001.
3. Parker, D.; Vieira, R. *Priorities for Energy Efficiency in New Residential Construction in Florida*; Florida Solar Energy Center: Cocoa, Florida, 2000; February.
4. Parker, D.; Sherwin, J. Comparative Summer Attic Thermal Performance of Six Roof Contructions, The 1998 ASHRAE Annual Meeting, Toronto, Canada, June 20–24, 1998.
5. Parker, D.; Sherwin, J.; Sonne, J.; Barkaszi, S.; Floyd, D.; Withers, C. Measured Energy Savings of a Comprehensive Retrofit in an Existing Florida Residence, For the Florida Energy Office, December 1997.
6. Fairey, P. Designing and Installing Radiant Barrier Systems, FSEC-DN-7, Florida Solar Energy Center, Cape Canaveral, FL, 1984.
7. Fairey, P. Effects of Infrared Radiation Barriers on the Effective Thermal Resistance of Building Envelopes, Proceedings of the ASHRAE/DOE Conference on Thermal Performance of the Exterior Envelopes of Buildings II, Las Vegas, NV, December 1982.
8. Fairey, P. The Measured Side-by-Side Performance of Attic Radiant Barrier Systems in Hot-Humid Climates, Proceedings of the 19th International Thermal Conductivity Conference, Cookeville, TN, October 1985.
9. Fairey, P. Radiant Energy Transfer and Radiant Barrier Systems in Buildings, FSEC-DN-6, Florida Solar Energy Center, Cape Canaveral, FL, 1984.
10. Joy, F.A. Improving attic space insulating values. ASHAE Trans. **1958**, *64*.
11. Levins, W.P.; Karnitz, M.A. Cooling-Energy Measurements of Unoccupied Single-Family Houses with Attics Containing Radiant Barriers, Oak Ridge National Laboratory, Contract Report, DE-ACO5-840R21400, July 1986.
12. Radiant Barriers: How They Work and How to Install Them, videotape, FSEC Producer, Cape Canaveral, FL, 1986.
13. Van Stratten, J.F. *Thermal Performance of Buildings*; Elsevier Publishing: New York, 1967.

Reciprocating Engines: Diesel and Gas

Rolf D. Reitz

Kevin Hoag

Engine Research Center, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.

Abstract

This entry first identifies the energy conversion principles that govern reciprocating piston devices. Next, the hardware and operating cycles used in practical reciprocating piston internal combustion engines are described. Finally, current issues, trends, and analysis approaches used in developing combustion and air handling systems are discussed.

INTRODUCTION

The goal of any engine is to convert the energy contained in a fuel into useful work as efficiently and cost-effectively as possible. Engines can be classified into various families based on how this conversion is accomplished. The internal combustion engine is one such family; in this family, a combustion reaction between a fuel and oxygen occurs within a mechanical device designed to harness the energy and produce work. This discussion addresses one mechanical configuration in the family of internal combustion engines. Other mechanical configurations of practical internal combustion engines are the gas turbine and rotary (Wankel) engine.

ENERGY CONVERSION AND RECIPROCATING PISTON DEVICES

Consider a sealed chamber in which a combustion reaction has just occurred, as shown in Fig. 1A. When a fuel reacts with oxygen, the chemical energy contained in the fuel is converted to sensible energy, which is physically sensed as a temperature increase in an open flame, and if the chamber is confined as shown here, an increase in temperature and pressure. The state within the chamber after combustion is indicated on coordinates of pressure and specific volume in the figure.

One can now begin considering various approaches for energy conversion. The possibilities include:

- Heat transfer from the chamber supplying energy to another device
- Mass transfer from the chamber and flow through a turbine
- Work transfer across a moveable chamber wall

Keywords: Thermodynamics; Piston engines; Energy conversion; Internal combustion engines; Spark ignition; Compression ignition.

Each possible approach follows some process line, as shown in Fig. 1B, which will ultimately bring the combustion gas to equilibrium with the outside environment. At equilibrium, the gas will have the same pressure, temperature, and chemical make-up as its surroundings. Unfortunately, any practical engine will release the gas to the environment before equilibrium is completely attained.[1] In other words, the temperature, pressure, or mixture concentration will still differ from the environment, but it becomes impractical to recover any further work.

If the curves shown were extended sufficiently to the right, each of these processes would have a different temperature when reaching atmospheric pressure. If work is being produced throughout the chosen process, intuition suggests that the process with the lowest temperature when atmospheric pressure is reached will be most efficient. The ideal process would simultaneously reach both pressure and temperature equilibrium with the environment. However, according to the Second Law of Thermodynamics, such a process does not exist.[1] The very best process—the highest efficiency process—is the one in which constant entropy is maintained. The temperature of the gas when atmospheric pressure is reached is still above atmospheric temperature, but all observations have demonstrated that a more efficient process cannot be achieved.

The reciprocating piston engine is one in which work transfer occurs across a moving combustion chamber wall (piston), as depicted in Fig. 1C. The pressure in the chamber forces the piston to the right, and work is performed on the piston. An important attraction of this moving piston device is that it has been found to follow a process line very closely resembling the isentropic process.

A very important practical concern arises: expansion to atmospheric pressure requires moving far to the right on the pressure versus specific volume coordinates. Because there is no change in the gas mass, the specific volume scale can be replaced with a volume scale, and it must be

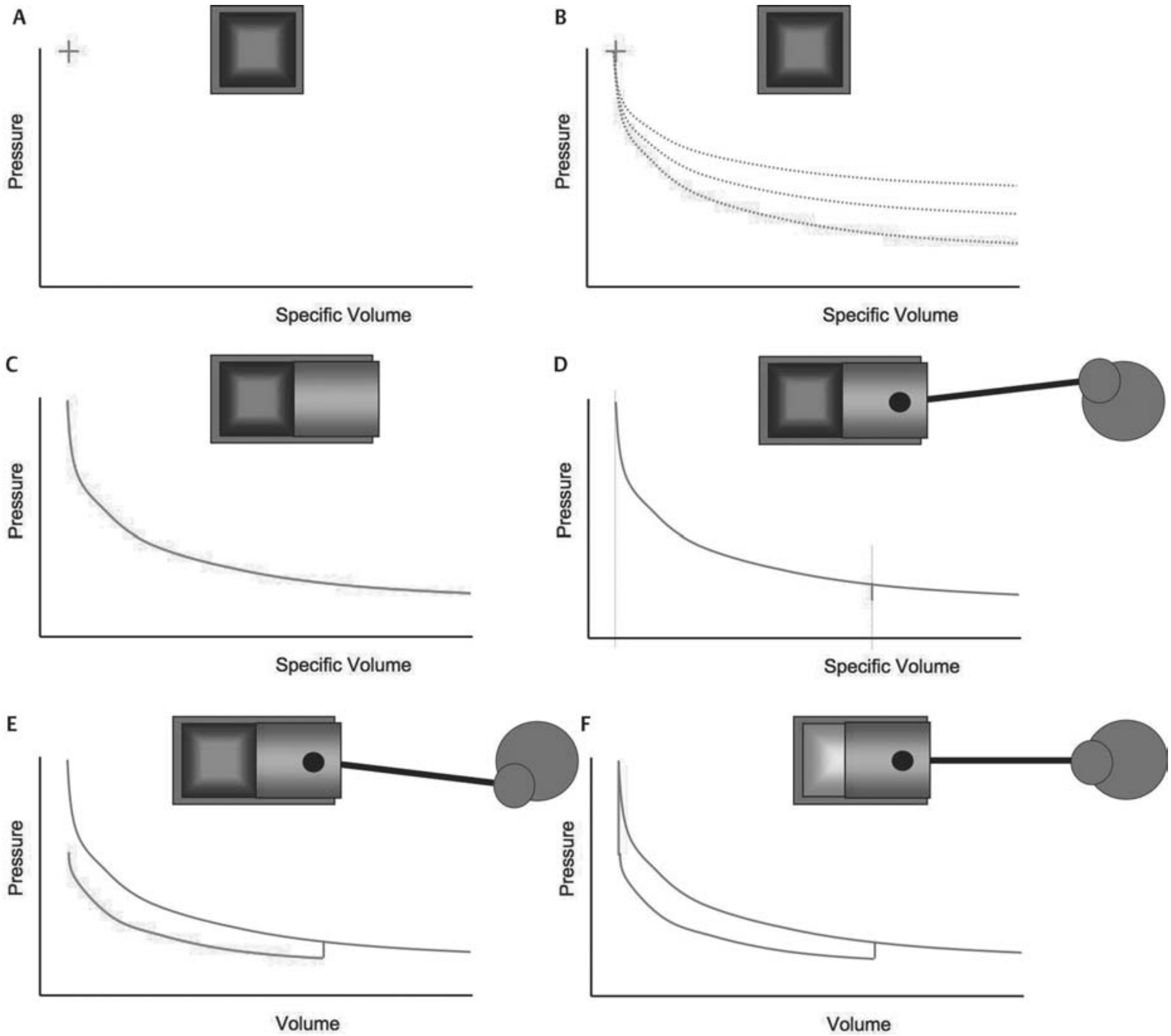


Fig. 1 Thermodynamic idealizations of reciprocating piston engine processes.

concluded that expansion to atmospheric pressure requires a very long chamber.

Fortunately, the isentropic process is quite non-linear. The pressure drops rapidly and the majority of the work extraction occurs early in the process. For practicality in regard to engine packaging, piston movement is stopped at some chosen maximum volume, and the remaining gas is expelled from the chamber.

In the arrangement shown in Fig. 1D, the piston is linked through a connecting rod to a crankshaft. This arrangement results in a defined maximum volume and the opportunity to create a repeated series of processes in which the piston is cycled between this maximum and some minimum volume. Thus, with suitable arrangements for bringing in a new mixture of fuel and oxygen and initiating combustion, work can be repeatedly produced. This arrangement also provides for convenient work extraction through the spinning crankshaft.

Fig. 1E shows a compression process. The linkage arrangement returns the piston to minimum chamber volume with a new charge in the cylinder. Work is required to compress the new charge, but here, too, the moving piston results in compression in a nearly 100% isentropic efficiency process.[1]

The cycle concludes as shown in Fig. 1F—with another combustion reaction from which the processes are repeated. This has been a conceptual discussion that demonstrates process idealizations. Much of engine development involves controlling the actual, practical processes to accomplish an engine operating cycle, and minimizing the deviations of the actual processes from the ideals. Practical processes for transferring fresh fuel and oxygen into the chamber, initiating and controlling the combustion process, and transferring the reacted gases out of the chamber are fundamental to the work of engine development.

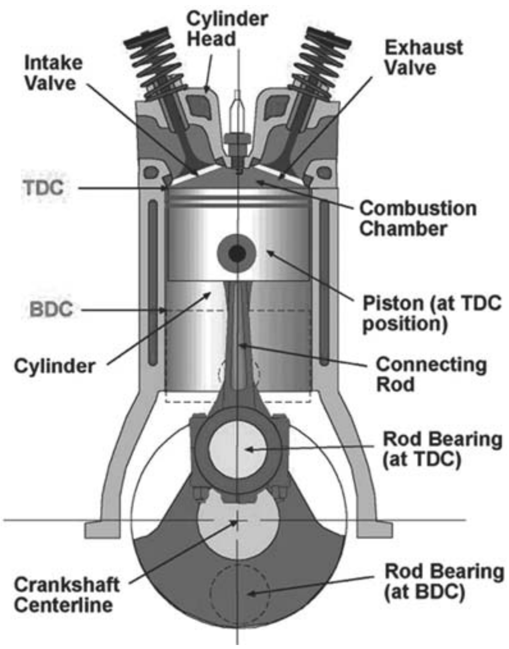


Fig. 2 Key components of reciprocating piston engines.

PRACTICAL ENGINE HARDWARE AND OPERATING CYCLES

The latter discussion identified the rationale behind the reciprocating piston engine. While many other mechanical configurations have been proposed, patented, and demonstrated over the years, few have enjoyed commercial success. Such success results from the ability to address both efficiency and cost-effectiveness.

The primary components of the reciprocating engine are shown in Fig. 2. The moving piston controls the volume of the combustion chamber between a minimum at top dead center (TDC) and a maximum at bottom dead center (BDC). The ratio between the volume at BDC and TDC is referred to as the compression ratio. The change in volume is the displacement of the cylinder, which, when multiplied by the number of cylinders, is the engine displacement.

The cylinder is sealed by the cylinder head opposite the moving piston. In most engines, the intake and exhaust valves are located in the cylinder head, as shown. The piston is linked to the crankshaft through a connecting rod. As the crankshaft spins about its centerline (the main bearing bore in the cylinder block), the offset of the rod bearing from the main bearing determines piston travel. As the crankshaft rotates a half revolution from the position shown in the figure, the piston moves from its TDC to its BDC position. The distance the piston travels is the engine's stroke, equal to twice the offset between the main bearing and rod bearing centerlines of the crankshaft. The cylinder diameter is termed its "bore"; the combination of bore and stroke determines the cylinder displacement.

Actual engine operation, based on the idealized operating cycle, can be further explained with reference to the pressure-volume diagrams of Fig. 3. The diagram in Fig. 3A is comprised of a four-stroke engine, while the Fig. 3B diagram depicts a two-stroke engine. The four-stroke engine requires four strokes of the piston—or two revolutions of the crankshaft—for completion of each cycle. The additional revolution is used to exhaust spent combustion products on the upward stroke of the piston, and draw in fresh reactants on the downward stroke. These additional strokes result in the relatively low pressure "pumping" loop at the base of Fig. 3A. The four strokes can be summarized as follows:

- Intake—the piston moves from TDC to BDC with the intake valve open, drawing fresh reactants into the cylinder.
- Compression—the valves are closed and the piston moves from BDC to TDC, compressing the reactants. The combustion reaction is initiated as the piston approaches TDC, increasing the pressure and temperature in the sealed chamber.
- Expansion—the high pressure forces the piston from TDC to BDC, transferring work to the crankshaft.
- Exhaust—the exhaust valve opens and the piston moves from BDC to TDC, forcing the spent exhaust products out of the chamber.

In the two-stroke engine, the operating cycle is completed in a single crankshaft revolution. Gas exchange is completed as the piston approaches BDC. Intake and exhaust passages are simultaneously opened, and a pressurized intake system is used to force the exhaust products out and fill the cylinder with a fresh charge in this scavenging process. The scavenging process results in the small process loop at the right of Fig. 3B (near BDC volume).

Today, considerable effort is devoted to gas exchange or "breathing" processes in engine development. Design goals include minimizing flow restrictions, maximizing spent products removal, and maximizing cylinder-filling with fresh reactants. Note that conventional spark-ignited gasoline engines operate with a premixed, nearly homogeneous fuel–air mixture. Load is controlled by throttling because of the relatively narrow ignitability limits of gasoline-air-residual mixtures, and this incurs substantial pumping losses at part load.

Another important aspect of engine development pertains to the combustion process. The idealization of Fig. 1F assumes an instantaneous combustion reaction, while the piston is at TDC. While this would result in maximum efficiency, real combustion processes are rapid, but not instantaneous. This fact, and the need to manage the peak pressure seen by the engine's structure, result in

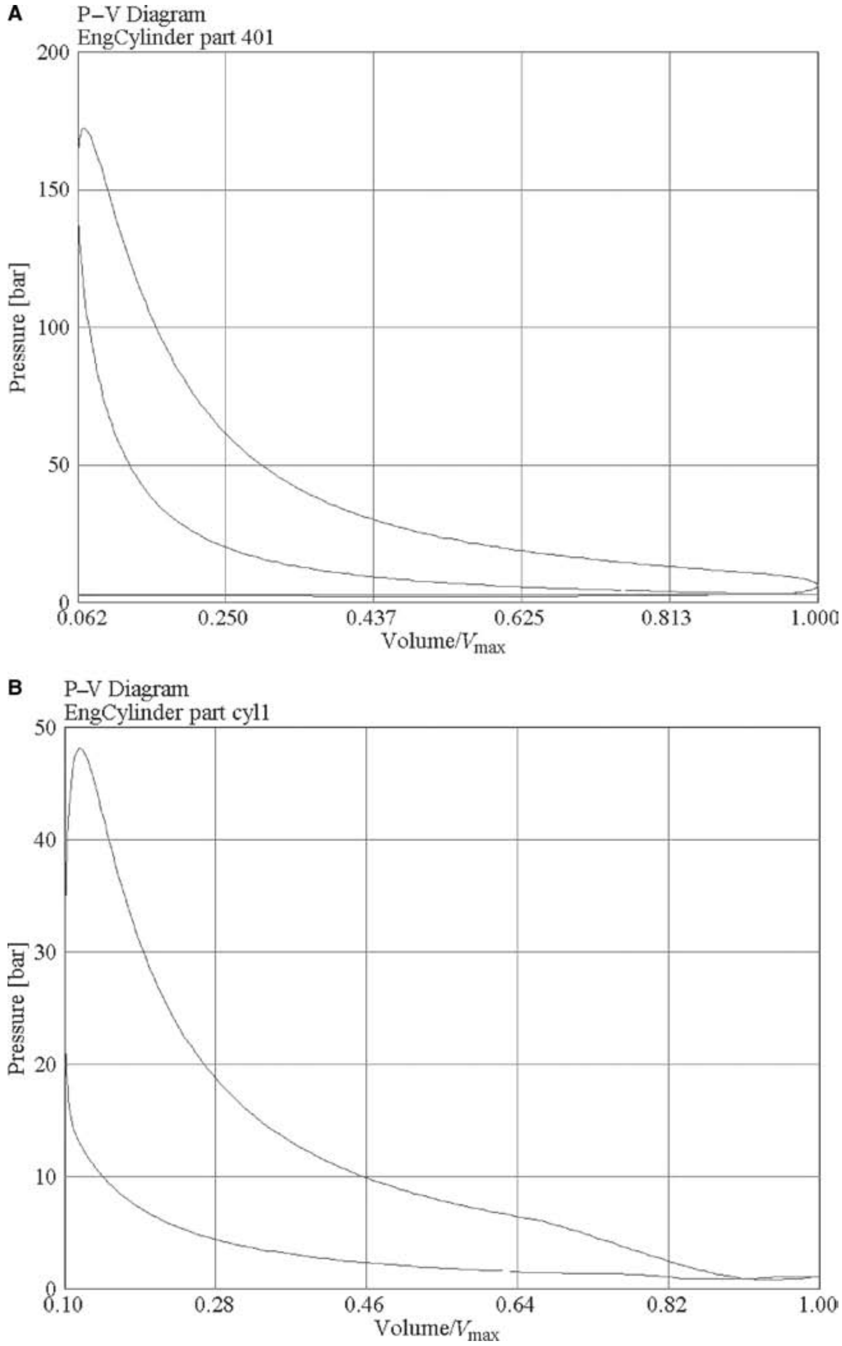


Fig. 3 (A) Four-stroke diesel pressure-volume diagram at full load. (B) Two-stroke spark-ignition pressure-volume diagram at full load.

combustion processes that are initiated slightly before TDC as the piston nears the end of the compression stroke. Combustion continues as the piston moves down during the expansion stroke.

Two approaches are prevalent in production engines. In a conventional spark-ignition engine, a mixture of air (the oxygen carrier) and fuel are drawn into the combustion chamber during the intake process. The mixture is compressed, and combustion is then initiated using a high-energy electrical spark. In the compression-ignition (diesel) engine, air alone is drawn into the cylinder and compressed. Fuel is injected directly into the cylinder near the end of the compression process. The fuel used in the compression-ignition engine is intended to easily spontaneously ignite when exposed to the temperature and pressure of the compressed air. While the diesel engine is often portrayed as having a slower combustion process (constant pressure instead of constant volume in the idealization of Fig. 1F), the goal of rapid combustion near TDC for maximum efficiency applies to both diesel and spark-ignition engines.

In order to increase the specific power output of an engine (power output per unit of displacement), some form of pre-compression (supercharging) is often considered. This is rapidly becoming standard practice in diesel engines, and is often seen in high-performance spark-ignition engines.

A crankshaft-driven compressor may be added to elevate the pressure of the air (or mixture) prior to drawing it into the cylinder. This allows more mixture to be burned in a given cylinder volume. Recognizing that the exhaust gases leaving the engine still contain a significant quantity of energy, an alternative is to use a portion of this energy to drive the compressor. This configuration is the turbocharged engine.

When air is compressed, its temperature increases. Its density can be further increased (and still more air forced into the cylinder) if it is cooled after compression. The charge air cooler, variously termed intercooler (cooling between stages of compression) and aftercooler (cooling after compression) may be used with either the turbocharger or crankshaft-driven supercharger.

ENGINE ANALYSIS

The remaining sections of this entry address analysis and development of practical engines, emphasizing engine breathing and combustion process optimization to meet efficiency and exhaust emission control needs. After a century of development it might be believed that the internal combustion engine would have little potential for further improvement. Nevertheless, new emissions standards have been promulgated that require the consideration of new technologies.[2] In response, engines continue to show substantial improvements in efficiency, power, and reduced emissions. Some of these advances have been due to the use of detailed analysis tools. Recent advances in computer speed and model development make the use of modeling increasingly attractive for engine design. Analysis tools include zero-dimensional thermodynamic models to analyze the in-cylinder combustion heat release, one-dimensional flow models to design the air handling and fuel injection systems, and multi-dimensional models that are useful in optimizing in-cylinder combustion and engine coolant systems.[3]

Gas Exchange Process

The gas exchange process controls the power achievable from the engine.[1] A conventional premixed-charge gasoline (i.e., spark-ignition) engine's intake system consists of the air filter, carburetor, and throttle plate or port fuel injector, intake manifold, intake port, and intake valves. Supercharging can be used to increase air mass into the cylinder in both gasoline and diesel engines. Spatial and temporal variations in flow and pressure throughout intake and exhaust manifolds must be considered to maximize cylinder filling and scavenging.

Intake system pressure drop occurs due to quasi-steady effects (flow resistances), and unsteady effects (wave action in the runners). Wave propagation is exploited in engine tuning and can be modeled effectively using commercial one-dimensional fluid flow software. Modeling is also needed to ensure equal airflows to each cylinder in a multi-cylinder engine. Engine breathing is affected by the intake and exhaust valve lifts and open areas (see Fig. 4), where much of the loss occurs, and valve overlap, which can cause exhaust gases to flow back into the intake system or intake gases to enter the exhaust (depending on the intake/exhaust pressure ratio), especially at low engine speeds.[4] At high engine speed, valve overlap can improve breathing since the inertia of the flowing gases can cause inflow into the cylinder even during the compression stroke. With Variable Valve Actuation (VVA) technologies, valve timing can be used to control the effective compression ratio (through early or late intake valve closure), or with exhaust gas re-induction (re-breathing) to control in-cylinder temperatures. Residual gas left from the previous cycle affects the engine combustion processes through its influence on charge mass, temperature, and dilution.

Accurate descriptions of valve flow losses require consideration of multidimensional flow separation

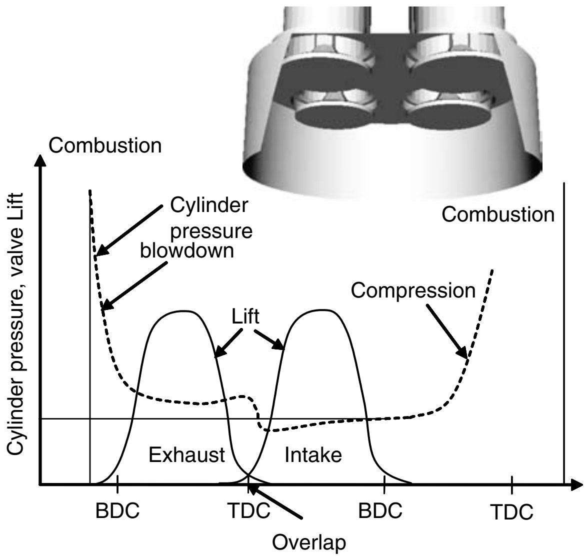


Fig. 4 Intake and exhaust valve arrangements for diesel engine.

Pum–Ren

Table 1 Heavy-duty diesel engine specifications

Engine details	Caterpillar 3401	Cummins N14
Cylinder bore (mm)	137.2	138.3
Stroke (mm)	165.1	152.4
Compression ratio	16.1	15.1
Displacement /cyl (l)	2.44	2.48
Fuel injection system		
Injectors	6 or 8 holes, 0.26 mm diameter	
Caterpillar port/ valve data		
Exhaust valve opening	−217° ATDC	
Intake valve opening	220° ATDC	
Intake valve closing	−143° ATDC	
Intake valve diameter	45 mm	
Exhaust valve diameter	41.8 mm	
Maximum valve lift	11.0 mm	
Intake port diameter	40.38 mm	
Exhaust port diameter	37.21 mm	
Port lengths	130 mm	

phenomena and their effect on the conditions within the cylinder at intake valve closure.[5] A typical valve timing diagram and predictions of the in-cylinder flow during the gas exchange process for a heavy-duty diesel engine (see Table 1) are shown in Figs. 4 and 5, respectively.[6] Fig. 5 shows velocity vectors and residual gas mass distributions in the engine just as the intake valves are about to close. The highest mixing of incoming fresh charge and combustion products occurs where the intake flow velocities are the largest due to high flow turbulence.

DIESEL ENGINES

Improvements in diesel engine technology in the last decade have been due to factors such as the introduction of high-pressure electronic fuel injection systems and multistage turbocharging, guided by insights obtained from computer modeling. Simulating the diesel engine represents one of the most complex and comprehensive modeling problems conceivable. The Navier–Stokes equations must be solved with consideration of when moving boundaries, turbulence, heat transfer (including radiation), sprays, combustion and emissions, and commercial and open source computer codes[7] are available.

Computer models are not entirely predictive due to the wide range of length and time scales needed to describe engine fluid mechanics. For example, a modern truck diesel engine operates with injection pressures as high as 200 MPa, with injector nozzle holes less than 200 μm in diameter. Thus, the fuel jet enters the high-temperature combustion chamber gas close to sonic velocity (∼600 m/s) and breaks up into droplets with diameters less than 10 μm in microseconds. To begin to resolve these processes in a combustion chamber of bore 100 mm, more than $(10^4)^3 = 10^{12}$ grid points would be required. Current, practical computer storage capabilities are limited to about 10^5 or 10^6 grid points. Accordingly, sub-models are used to describe unresolved processes,[8] as summarized in Table 2.

The combustion chamber surface temperature is needed to specify wall boundary conditions (see Fig. 6[19]). In the model, heat flux from the gas-side is used for the metal component heat-conduction prediction. Localized high temperature regions are predicted near the piston bowl rim in agreement with engine experimental data.

Following fuel injection and atomization, injected fuel droplets vaporize and mix with the compressed air. Auto-ignition takes place and the burning rapidly consumes the premixed mixture. Thereafter, the burning that occurs while the injection continues occurs via mixing-controlled or diffusion-type combustion. Fig. 7A shows predicted spray plumes and soot formation regions for a Caterpillar diesel engine (see Table 1). Good agreement between measured and predicted in-cylinder pressures and soot and NO_X emissions are shown in Figs. 7B and 8, respectively.[5,12]

Fig. 8 also shows that staged or multiple injections are effective for emissions reduction, and the mechanism of emission reduction is summarized in Fig. 9.[20] Soot accumulates at the tip of the spray jet (see also Fig. 7A), as also confirmed experimentally in optical engines.[21] With single injection, the high-momentum injected fuel penetrates to the fuel-rich, relatively low-temperature jet tip and continuously replenishes it, producing soot. With a split-injection, the soot cloud of the first plume is not replenished with fresh fuel, but continues to oxidize.

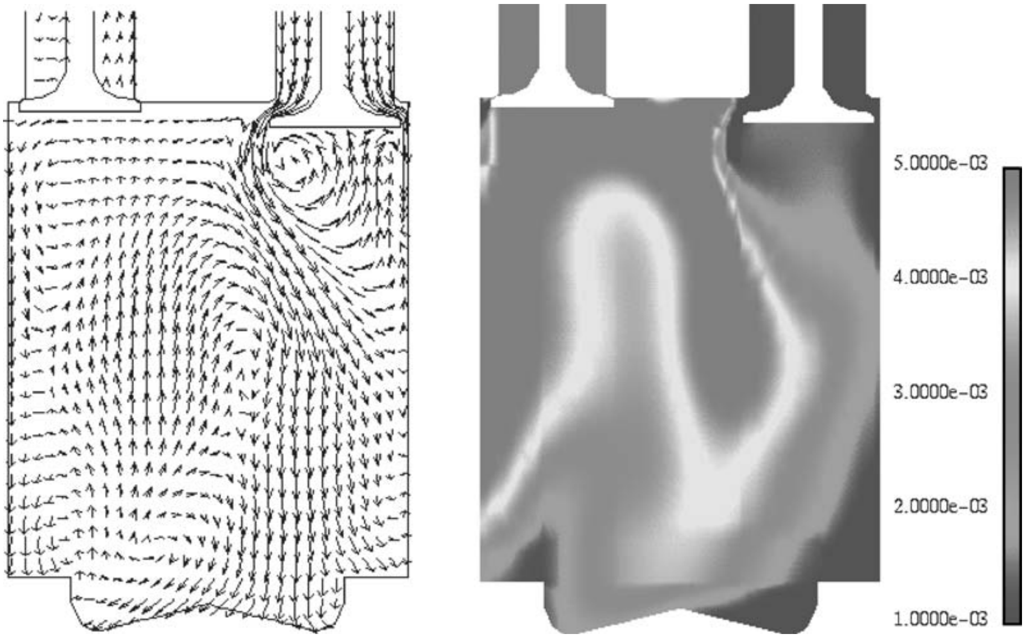


Fig. 5 Computed flow velocities (left) and residual gas distribution (right) during gas exchange in plane of valves (144 degrees ATDC—1600 rev/min)
Source: Reprinted with permission from SAE Paper 962052 © 1996 SAE International.

GASOLINE ENGINES

Gaseous pollutant emissions (NO_X, CO, and HC) from gasoline engines have been controlled effectively using the three-way catalyst. However, in order to permit both reductions of NO_X and the oxidation of hydrocarbons and CO, combustion must occur at the chemically balanced (stoichiometric) air–fuel ratio within very narrow limits. Issues such as cold-start emissions, fuel economy, and engine responsiveness have been the main drivers for further advancements. The fuel system of spark-ignition engines has evolved from carburetion to throttle-body injection, followed by Port Fuel Injection (PFI), including sequential-fire injection.

Table 2 Example physical submodels used in diesel engine computer modeling

Physical process	Submodel
Turbulence	RNG k-ε [9]
Heat transfer	Log-law [9]
Nozzle flow	Cavitation [10]
Atomization	Jet instability [11]
Drop breakup/coalesence	Drop instability [7,12]
Drop distortion	Drop drag [12]
Hydrocarbon ignition	Reduced [13]/detailed chemistry [14]
Diesel combustion	Turbulent/chemistry timescale [15,16]
NOx emissions	Zeldo'vich/detailed [15]
Soot emissions	Formation/Oxidation [16–18]

During cold start, a film of liquid fuel can form in the intake valve area of the port and some portion of it is drawn into the cylinder during induction, as shown in Fig. 10A. The fuel delivered to each cylinder can also differ from that metered by the injector, causing a fuel delivery delay and an associated metering error. Thus, extra fuel may be needed for cold starting, causing increased unburned hydrocarbons emissions.

With Gasoline Direct Injection (GDI), which is also called Direct Injection Spark Ignition (DISI), fuel is injected directly into the combustion chamber (Fig. 10B). This leads to higher fuel efficiency, better drivability, and better cold-start performance.[25] However, GDI engines suffer from higher hydrocarbon emissions at light-load, hampering introduction in some markets.

There are three GDI operation régimes.[25] The simplest use early injection to create a homogeneous stoichiometric mixture, and throttling for load control. Cooling due to vaporization of injected fuel allows for the use of higher compression ratios and lower octane fuels without knock (uncontrolled fast combustion). The next régime uses a lean homogeneous mixture with reduced throttling for greater load control. In practice, lean homogeneous operation is less desirable than heavy

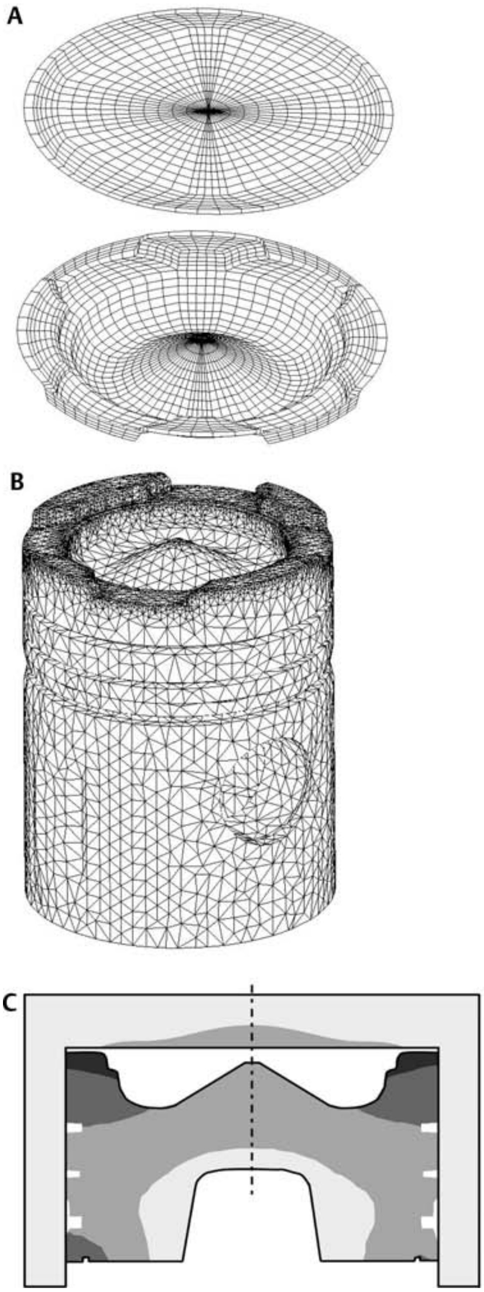


Fig. 6 Cummins diesel engine (A) in-cylinder CFD surface mesh, (B) finite element piston mesh, (C) predicted temperature distribution on head, liner, and piston.
Source: Reprinted with permission from SAE Paper 2003-01-0560 © 2003 SAE International. (see Ref. 19).

dilution with Exhaust Gas Recirculation (EGR) to reduce throttling since overall lean operation precludes the use of a three-way catalyst and requires more elaborate lean-NO_X after-treatment approaches, such as storage catalysts.

The greatest benefit of GDI is operation in the third regime, which is GDI sans throttling with an overall lean stratified mixture at part load and late-injection. This strategy should smoothly transition to homogeneous charge operation by injecting a larger mass of fuel early in the cycle. Engines utilizing all three GDI modes require complex control systems.

Accurate fuel delivery makes it possible to operate with more dilute mixtures and less cycle-to-cycle variation. Visualization experiments have provided substantial insights into mixture preparation and the causes of cyclic variability in ignition and combustion.[25,26] An inherent disadvantage of lean and stratified GDI engines is that with light-load operation, a three-way catalyst cannot be used effectively for NO_X control due to excess air in the exhaust stream. In addition, since the location of the ignition source is fixed, the mixture cloud must be controlled both temporally and spatially for optimal spark ignition, requiring precise matching of fuel injection and in-cylinder flow fields. CFD modeling has been useful for design,[24] and spark-ignition kernel growth and flame propagation models based on front tracking techniques, or flamelet models have been developed for the simulations.[27]

Two mixing concepts have been proposed: the "wide spacing" concept,[28] where the injector is side-mounted and fuel is injected toward the piston surface and deflected toward the spark plug (see Fig. 10C).[24] Spray momentum is independent of engine speed, and is the driving force for air–fuel mixing; adequate mixing can thus be realized over a wide range of engine speeds. The "close spacing" or "spray-guided" concept, with central injector placement (see Fig. 10B), is often seen in homogenous charge GDI engines,[23] but is also under development as a means to increase the speed-load range over which highly stratified operation can be maintained.[29] To improve the transition between low load and high load operation, optimally timed split injections during intake and compression have been proposed.[30]

HCCI ENGINES

There is great current interest in Homogeneous Charge, Compression Ignition (HCCI) engines for both light-duty automotive (gasoline and diesel) and heavy-duty diesel engines due to its potential for very low emissions and high fuel efficiency. In these engines, a mixture of fuel, air, and recycled combustion products is compressed until it auto-ignites.[31] Unlike diesel engines, combustion is not controlled by the fuel injection, and unlike spark-ignited engines, no discernable flame front is evident. At light load, low temperature combustion leads to more than 90% lower NO_X emissions compared to conventional diesel engines.[32] Low soot emissions are also generally found. Unburned hydrocarbon emissions can increase with fuel deposition or lean operation (at low loads), but can be

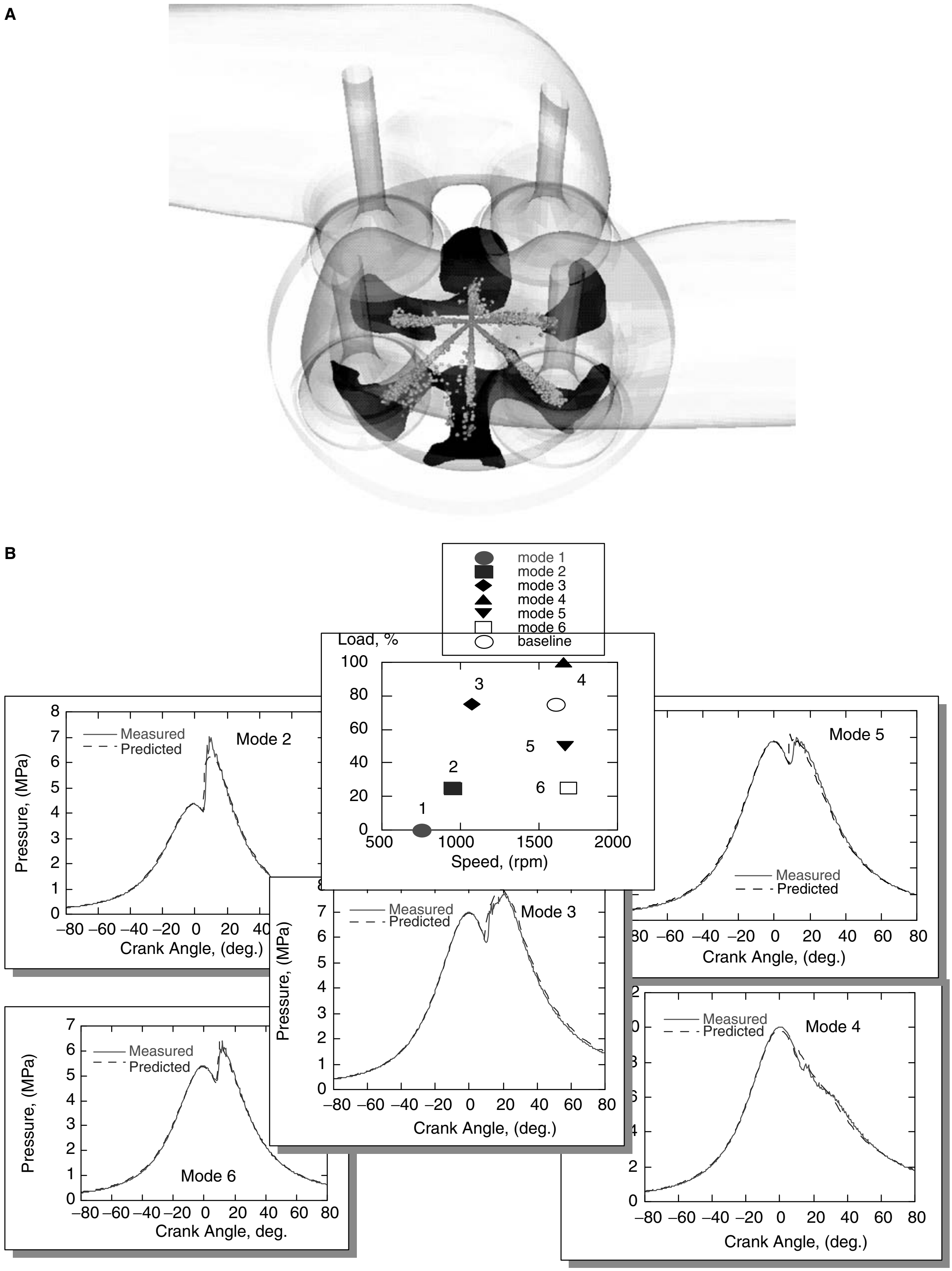


Fig. 7 (A) Fuel sprays with soot iso-surfaces (dark region at spray tips) during combustion, (B) measured and predicted cylinder pressure at steady-state test points for a heavy-duty diesel engine.
Source: Reprinted with permission from SAE Paper 2000-01-1178 © 2000 SAE International.

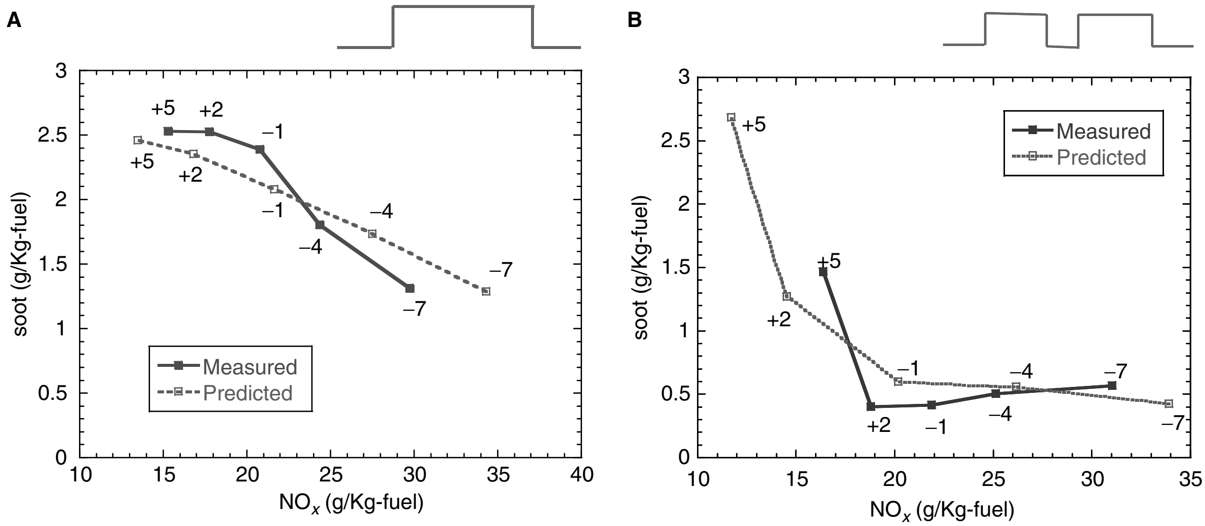


Fig. 8 Measured and predicted soot-NOx tradeoff for a heavy-duty diesel engine. (A) single and (B) double injection (1600 rev/min, 75% load). Numbers indicate start-of-injection timing in crank-angle degrees, where negative numbers are before TDC in the compression stroke and positive numbers are after TDC.
Source: Reprinted with permission from SAE Paper 2000-01-1178 © 2000 SAE International (see Ref. 5).

reduced with available exhaust gas oxidation catalysts. Fuel impingement can be reduced with low pressure, low penetration sprays. The requirement to use lean air–fuel mixtures limits the achievable power, favoring the use of high boost.

Challenges to HCCI operation include the fact that there is no direct method for controlling the start-of-combustion timing, leading to poor fuel efficiency. Ignition is controlled by chemical kinetics, and influenced by the fuel composition, mixture stoichiometry, and the thermodynamic state of the mixture. There are no external controls, such as the fuel injection or spark timings used in diesel or SI engines, respectively. However, charge stratification using multiple injection strategies has been explored for combustion phasing and heat release rate control.[33] Due to the rapid auto-ignition heat release, true HCCI combustion approaches the idealization described in the first section. This rapid heat release can cause engine noise and potentially damage the engine, and this places constraints on the maximum load or mixture strengths that can be used. Diesel ignition delay can be extended to allow time for mixing by using very early injection or late injection, as in Nissan's Modulated Kinetics (MK) concept.[34]

CONCLUSIONS

This entry provides a review of diesel and gasoline engine operating principles, and an introduction to current research issues and literature. Developments in engine hardware, coupled with advanced analysis techniques, such as computer modeling, allow efficient exploration of the large number of possible design configurations. As an

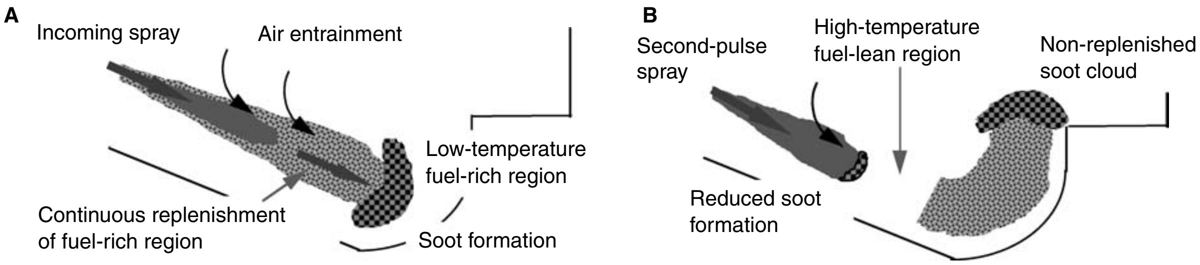


Fig. 9 Mechanism of emission reduction with multiple injections.
Source: Reprinted with permission from SAE Paper 960633 © 1996 SAE International (see Ref. 20). (A) single injection, (B) split injection.

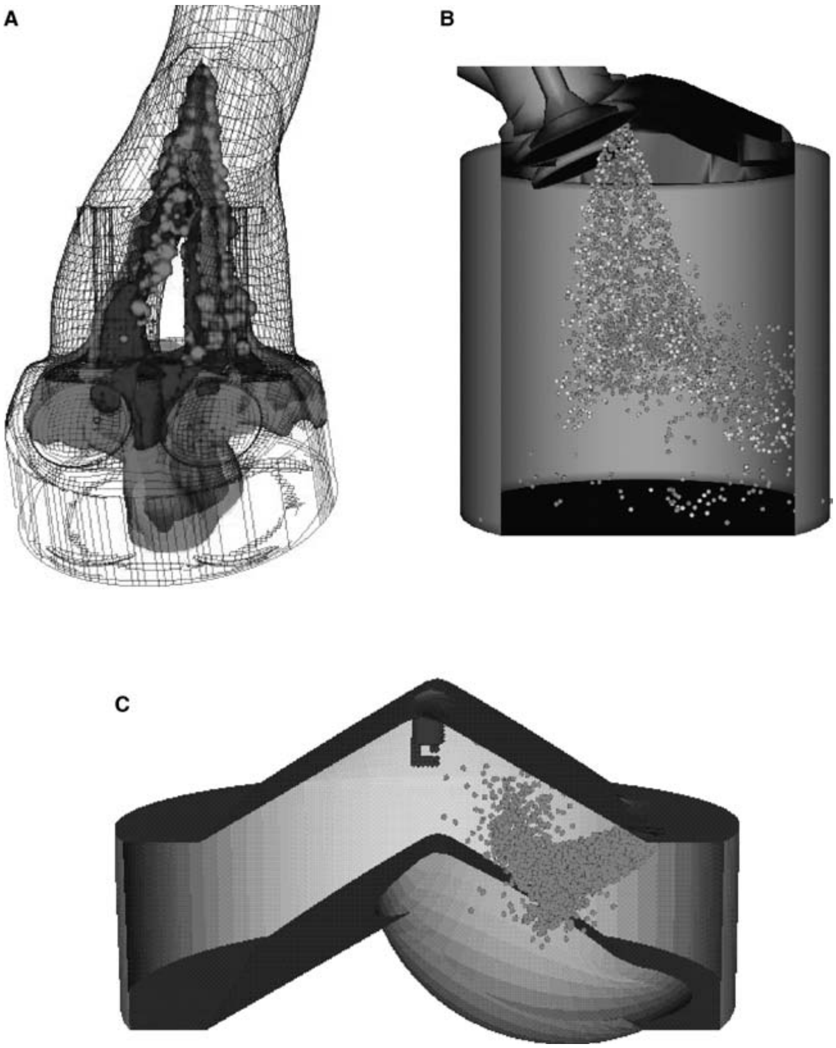


Fig. 10 Gasoline engine fuel–air mixing (A) Port Fuel Injection (PFI).
Source: Reprinted with permission from SAE Papers 2001-01-1231 © 2001(see Ref. 22), (B) spray-guided direct injection.
Source: Reprinted with permission from SAE Papers 970625 © 1997 SAE International (see Ref. 23), (C) wall-guided direct injection.

example of current technology, the combustion chamber geometry can be optimized using computer modeling for low emission, high efficiency engine operation.[35]

ACKNOWLEDGMENTS

Contributions of faculty, students, and staff at the Engine Research Center are greatly appreciated.

REFERENCES

1. Heywood, J.B. *Internal Combustion Engine Fundamentals*; McGraw-Hill Inc.: New York, 1988.
2. Anon, Emission Standards and Supplemental Requirements for 2007 and Later Model Year Diesel Heavy-Duty Engines and Vehicles, 40 CF—Chapter I—Part 86, 40CFR68.007-11, US EPA, March, 2002.
3. Reitz, R.D. Computational fluid dynamics modeling of diesel engine spray combustion and emissions. In *Diesel Engine Reference Book*, 2nd Ed.; Challen, B., Baranescu, R., Eds.; Butterworth-Heinemann: Woburn, MA, 1998.
4. Zhu, Y.; Reitz, R.D. A 1-D Gas Dynamics Code for Subsonic and Supersonic Flows applied to predict EGR levels in a heavy-duty diesel engine. Int. J. Vehicle Design **1999**, *22*, 227–252.
5. Yi, Y.; Hessel, R.; Zhu, G.; and Reitz, R.; The Influence of Physical Input Parameter Uncertainties on Multidimensional Model Predictions of Diesel Engine Performance and Emissions, SAE Paper 2000-01-1178, SAE Transactions, Engines, Vol. 109, Section 3, pp. 1298-1316, 2000.
6. Senecal, P.K.; Xin, J.; Reitz, R.D.; Predictions of Residual Gas Fraction in IC Engines, SAE Paper 962052, SAE Transactions, Vol. 105, Section 3, J. Engines, pp. 2243-2254, 1996.

7. Amsden, A.A.; KIVA-3V: A Block-Structured KIVA Program for Engines with Vertical or Canted Valves, Los Alamos National Laboratory Report, LA-13313-MS (1997).
8. Reitz, R.D.; Rutland, C.J. Development and testing of diesel engine CFD models. Progr. Energy Combust. Sci. **1995**, *21*, 173–196.
9. Han, Z.; Reitz, R.D. Turbulence modeling of internal combustion engines using RNG k-e models. Combust. Sci. Tech. **1995**, *106*, 267–276.
10. Von Kuensberg, S.C.; Kong, S.-C.; Reitz, R.D.; Modeling the Effects of Injector Nozzle Geometry on Diesel Sprays, SAE Paper 1999-01-0912, 1999.
11. Reitz, R.D. Modeling atomization processes in high-pressure vaporizing sprays. Atom. Spray Tech. **1988**, *3*, 309–337.
12. Xin, J.; Ricart, L.; Reitz, R.D. Computer modeling of diesel spray atomization and combustion. Combust. Sci. Tech. **1998**, *137*, 171–183.
13. Halstead, M.; Kirsh, L.; Quinn, C. Multidimensional autoignition of hydrocarbon fuels at high temperatures and pressures—fitting of a mathematical model. Combust. Flame **1977**, *30*, 45–60.
14. Westbrook, C.K.; Warnatz, J.; Pitz, W.J.; A Detailed Chemical Kinetic Reaction Mechanism for the Oxidation of *Iso-Octane* and *N*-Heptane Over an Extended Temperature Range and its Application to Analysis of Engine Knock, Proceedings of 22nd Symposium (Int.) on Combust, pp. 893–901, Combustion Institute, Pittsburg, PA, 1988.
15. Kong, S.-C.; Patel, A.; Yin, Q.; Klingbeil, A.; Reitz, R.D.; Numerical Modeling of Diesel Engine Combustion and Emissions Under HCCI-Like Conditions with High EGR Levels, SAE Paper 2003-01-1087, SAE Transactions, Volume 112, Section 3, J. Engines, pp. 1500–1510, 2003.
16. Kong, S.-C.; Sun, Y.; Reitz, R.D.; Modeling Diesel Spray Flame Lift-Off, Sooting Tendency and NO_x Emissions Using Detailed Chemistry with Phenomenological Soot Model, Proceedings of ASME Spring Technical Conference, (ICES05), McCormick Place Convention Center, Chicago, April 5–7, 2005.
17. Hiroyasu, H.; Nishida, K.; Simplified Three-Dimensional Modeling of Mixture Formation and Combustion in a D.I. Diesel Engine, SAE Paper 890269, 1989.
18. Nagel, J.; Strickland-Constable, R.F.; Oxidation of Carbon Between 1000-2000 C, Proceedings of the Fifth Carbon Conference; Pergammon Press: Elmsford, NY, Vol. 1, 1962; pp. 154–172.
19. Wiedenhoefer, J.F.; Reitz, R.D.; Multidimensional Modeling of the Effects of Radiation and Soot Deposition in Heavy-Duty Diesel Engines, SAE Paper 2003-01-0560, SAE Transactions, Volume 112, Section 3, J. Engines, pp. 784–804, 2003.
20. Han, Z.; Uludogan, A.; Hampson, G.; Reitz, R.D.; Mechanisms of Soot and NO_x Emission Reduction Using Multiple-Injection in a Diesel Engine, SAE Paper 960633, SAE Transactions, Vol. 105, Section 3, J. Engines, pp. 837–852, 1996.
21. Dec, J.E.; A Conceptual Model of DI Diesel Combustion Based on Laser-Sheet Imaging, SAE Paper 970873, 1997.
22. Zhu, G.-S.; Reitz, R.D.; Xin, J.; Takabayashi, T.; Characteristics of Vaporizing Continuous Multi-Component Fuel Sprays in a Port Fuel Injection Gasoline Engine, SAE Paper 2001-01-1231, SAE Transactions, Vol. 110, J. Engines, Section 3, pp. 1368–1384, 2001.
23. Han, Z.; Reitz, R.D.; Yang, J.; Anderson, R.W.; Effects of Injection Timing on Air-Fuel Mixing in a Direct-Injection Spark-Ignition Engine, SAE Paper 970625, SAE Transactions, Vol. 106, Section 3, J. Engines, pp. 848–860, 1997.
24. Fan, L.; Reitz, R.D. Spray and combustion modeling in gasoline direct-injection engines. Atom. Sprays **2000**, *10* (3–5), 219–250.
25. Zhao, F.Q.; Lai, M.C.; Harrington, D.L.; A Review of Mixture Preparation and Combustion Strategies for Spark Ignited Direct Injection Gasoline Engines, SAE Paper 970627.
26. Fansler, T.D.; Drake, M.C.; Stojkovic, B.; Rosalik, M.E. Local Fuel Concentration, Ignition and Combustion in a Stratified-charge spark-ignited direct-injection engine: Spectroscopic, imaging and pressure-based measurements. Intl. J. Engine Res. **2003**, *4*, 61–86.
27. Tan, Z.; Reitz, R.D.; Ignition and Combustion Modeling in Spark-Ignition Engines Using a Level Set Method, SAE Paper 2003-01-0722, SAE Transactions, Volume 112, Section 3, J. Engines, pp. 1028–1040, 2003.
28. Iwamoto, Y.; Noma, K.; Nakayama, O.; Yamauchi, H.; Ando, H.; Development of Gasoline Direct Injection Engines, SAE Paper 960627.
29. Drake, M.C.; Fansler, T.D.; Lippert, A.M. Stratified-charge combustion: Modeling and imaging of a spray-guided direct injection spark-ignition engine. Proc. Combust. Inst. **2005**, *30*, 2683–2691.
30. Stiesch, G.; Tan, Z.; Merker, G.P.; Reitz, R.D.; Modeling the Effect of Split Injections on Direct Injection Spark Ignition (DISI) Engine Performance, SAE Paper 2001-01-0965, 2001.
31. Najt, P.; Foster, D.; Compression-Ignited Homogeneous Charge Combustion, SAE Technical Paper, No. 830246, 1983.
32. Stanglmaier, R.H.; Roberts, C.E.; Homogeneous Charge Compression Ignition (HCCI): Benefits, Compromises, and Future Engine Applications, SAE Paper 1999-01-3682.
33. Marriott, C.M.; Reitz, R.D.; Experimental Investigation of Direct Injection-Gasoline for Premixed Compression-Ignited Combustion Phasing Control, SAE Paper 2002-01-0418, SAE Transactions, Volume 111, Section 3, J. Engines, pp. 863–874, 2002.
34. Kimura, S.; Aoki, O.; Kitahara, Y.; Aiyoshizawa, E.; Ultra-Clean Combustion Technology Combining a Low-Temperature and Premixed Combustion Concept for Meeting Future Emission Standards, SAE Paper 2001-01-0200, 2001.
35. Wickman, D.D.; Yun, H.; Reitz, R.D.; Optimized Split-Spray Piston Geometry for HSDI Diesel Engine Combustion, SAE Paper 2003-01-0348, 2003.

Regulation: Price Cap and Revenue Cap

Mark A. Jamison
Public Utility Research Center, University of Florida, Gainesville, Florida, U.S.A.

Abstract
Price cap regulation allows the operator to change its price level according to an index that is typically comprised of an inflation measure, I, and a "productivity offset," which is more commonly called the X-factor. Typically with price cap regulation, the regulator groups services into price or service baskets and establishes an I–X index, called a price cap index, for each basket. Establishing price baskets allows the operator to change prices within the basket as the operator sees fit as long as the average percentage change in prices for the services in the basket does not exceed the price cap index for the basket. Revenue cap regulation is similar to price cap regulation in that the regulator establishes an I–X index, which in this case is called a revenue cap index, for service baskets and allows the operator to change prices within the basket so long as the percentage change in revenue does not exceed the revenue cap index. Revenue cap regulation is more appropriate than price cap regulation when costs do not vary appreciably with units of sales.

INTRODUCTION

Price cap and revenue cap regulation are forms of incentive regulation, which is the use of rewards and penalties to induce the utility company to achieve desired goals and in which the operator is afforded some discretion in achieving goals.[1,2] With price cap regulation, the company's average price increase is restricted by a price index that generally includes an inflation measure (such as the U.S. Gross Domestic Product Implicit Price Deflator) and an offset that generally reflects expected changes in the company's productivity. (Revenue cap regulation is the same as price cap regulation except that the company's revenue is restricted by the inflation-productivity index. In this chapter, I simplify my discussion by focusing on price cap regulation.) With *pure price caps*, the regulator never directly observes the operator's profits. This form of price caps is rare and indeed may never be practiced except in instances where the regulator is prohibited by law from observing costs and adjusting prices. Most price cap regimes base prices on past costs or expected costs, and prohibit the regulator from adjusting prices according to new information for a set period—typically, 4–6 year.

Price caps were first developed in the United Kingdom in the 1980s to be the regulatory framework for the country's newly privatized utilities. The basic idea behind the country's price cap regulation was that the regulator would be at an information disadvantage relative to the utilities in terms of knowing how efficiently the utilities could operate. By adopting price cap regulation and allowing utilities to keep for a period of time profits they received by improving efficiency, the government believed that the companies would reveal their efficiency capabilities. In turn, this would allow the regulator eventually to set regulated prices that reflected the companies' true abilities. Price cap regulation did not work out entirely as planned, so adjustments have been made to the point that the United Kingdom's price cap regulation looks a lot like U.S. rate of return regulation. (Excellent summaries of the U.K. experience can be found in several studies.[3–5] A critical difference between U.S.-style rate of return regulation and U.K.-style price cap regulation is that the U.K. regimes have fixed time periods between price reviews, whereas under rate of return regulation, price reviews are triggered by high or low earnings [relative to the cost of capital].)

There are three important elements of an incentive regulation plan: (1) providing the reward/penalty structure; (2) allowing the company an opportunity to choose its goals; and (3) allowing the operator latitude in how it will achieve its goals. An example of a reward/penalty structure would be allowing the company to retain higher (lower) profits if it increases (decreases) its operating efficiency. Allowing the company a role in choosing its goals is referred to as "a menu of options," whereby the regulator matches greater potential rewards with more ambitious goals. The company may be allowed to choose, for example, between a goal of decreasing costs by 5% and keeping 50% of the profits it receives above its cost of capital, and a goal of decreasing costs by 10% and keeping 100% of the profits it receives above its cost of capital. (A company's cost of capital is the interest that the company pays on its debt plus the return that it must provide shareholders to ensure that they continue to invest in the company.) If the company chose the goal of decreasing costs by 10%, the operator would have the latitude to do this by, for example, negotiating lower input

Keywords: Price caps; Incentives; Revenue caps; Information asymmetry.

prices from suppliers, decreasing overhead, improving network reliability, obtaining lower-cost capital, or using some combination of methods.

The benefits of price cap regulation include providing companies incentives to improve efficiency, dampening the effects of cost information asymmetries between companies and regulators, and decreasing the incentives to overinvest in capital and cross-subsidize relative to rate of return regulation. In some instances, however, service quality and infrastructure development have suffered under price cap regulation. Furthermore, it is difficult for regulators to keep commitments that allow companies to retain profits above their cost of capital.

The remainder of this chapter is organized as follows. The next section describes the theory underlying price cap regulation. "The Basic Price Restriction" describes establishing the price index. "Service Baskets" discusses how regulators structure price baskets. "Case Studies in Price Caps" summarizes some cases followed by the "Conclusion."

UNDERLYING THEORY

Regulators and other policymakers have certain energy goals for their countries, including near-universal availability of service, affordable prices, and quality service. Achieving these goals requires that utilities incur costs and exert effort. The difficult question for regulators is how much cost and effort will be required. Utilities generally know more about the answers to these questions than regulators do. A company generally knows more than its regulator about how much it would cost to provide a certain level and quality of network expansion, for example. This is because the regulator cannot directly observe the operator's innate abilities and its degree of effort.

These problems are called information asymmetry or principal-agent problems. An information asymmetry arises from the company's having information—namely, about the utility's innate ability to achieve performance goals at a specific cost and the amount of effort the employees exert—that the regulator does not have. The name "principal-agent" arises from the nature of the relationship; the regulator (the principal) has goals that she wants the operator (the agent) to achieve. The company may agree with some of the principal's goals, but companies generally have other interests, such as maximizing profits for their shareholders and limiting the amount of effort exerted. To solve these problems, the regulator offers the operator financial rewards for controlling costs and/or exerting effort.

THE BASIC PRICE RESTRICTION

With price cap regulation, prices are initially set to allow the company to receive its cost of capital. Thereafter, prices are allowed to rise on average at the rate of inflation, less an offset, namely

$$\%\Delta p \leq I - X,$$

where $\%\Delta p$ is the average percentage change in prices allowed in a year, I is the inflation index, and X is the offset. The key issues are: What is the "offset"? What is the measure of inflation? And what does it mean that prices are allowed to rise on average?[6]

The underlying logic of the price cap restriction is that it emulates the competitive market. In a competitive market, prices reflect the costs of production. Prices rise when production costs unavoidably rise. Prices decline with productivity increases. As a result, in a competitive economy, the economy-wide inflation rate reflects unavoidable increases in production costs and accounts for productivity gains. If the regulated company is just like the average firm in the economy, its prices should rise at the general rate of inflation.[7]

Therefore, the X-factor should represent the difference between the regulated firm and the average firm in the economy. There are two key differences to consider—namely, the regulated company's ability to improve productivity and changes in its input costs. If the regulated company can improve its productivity more than the average firm in the economy, or if the regulated company's input prices increase less than input prices for the average firm, this would imply $X>0$. The opposite situations would imply $X<0$. If the regulated firm is just like the average firm, this would imply $X=0$. Consider a situation in which the average firm in the economy improves its productivity by 3% per year, and its input prices increase 1% per year. Further assume that the regulated firm can improve its productivity by 5% per year and that its input prices actually decrease 2% per year. The appropriate X-factor would be $X = (5 - 3) - (-2 - 1) = 5$.

There are two basic approaches for establishing an X-factor—namely, the historical approach and the forecast approach. The historical approach compares estimates of the total factor productivity (TFP) for the average firm in the economy with estimates for the regulated company. The X-factor is set equal to the difference between the TFP estimates after adjusting for differences in input prices. A modification to this approach adds a stretch factor, S, that accounts for the effects of historic regulation and/or anticipated changes in industry conditions. Examples of explicit stretch factors include 0.5% for AT&T by the Federal Communications Commission and 1% for local-exchange telephone companies in Canada.

The forecast approach is a three-step process. The first step is to determine the rate base for year t, where t is the first year of the new pricing regime, according to the formula

$$B_{t-1} = B_0 + \sum_{i=1}^{t-1}(\text{Capex}_i - d_i),$$

where $t=0$ is the initial rate base of the company, for example, at the time of privatization; Capex_i is the additional investment in rate base in year i; and d_i is the depreciation expense in year i. The next step is to project cash outflows (Capex), operating expenses (Opex), non-operating expenses (Nopex), and unit sales for each year of the new pricing regime. The last step is to estimate the X-factor that will equate the present value of the cash flows of the company with the change in shareholder value, using the formula

$$\sum_{j=t}^{n+t} \frac{P_jQ_j - \text{Opex}_j - \text{Capex}_j - T_j \pm \text{Tr}}{(1 + \text{WACC})^j} = B_{t-1} - \frac{B_{n+t}}{(1 + \text{WACC})^n},$$

where P_jQ_j is the projected revenue for year j, Opex_j is the forecasted operating expenses for year j, Capex_j is the projected capital expenditures for year j, T_j is the projected taxes for year j, B_j is the rate base at end of year j, Tr represents cash transfers between the government and other entities (not counted in revenue, operating expenses, or capital expenditures), WACC is the weighted average cost of capital, and n is the length of time a price cap plan is in effect.[8] The WACC is the return the company is allowed to receive on its assets, and includes both the cost of debt the company uses to finance its rate base and the cost of equity. The cost of debt is simply the weighted average of the interest rates that the company pays on its long-term corporate bonds. The cost of equity is the return that shareholders need to ensure that they continue to finance the company.[9]

The United Kingdom used this approach in setting prices for Hydro Electric (HE) in 1995.[10] Table 1 shows the Monopoly and Merger Commission's (MMC) present value calculation for HE's price control for the period 1995–1996 to 1999–2000. The first three lines contain its allowances for operating costs, network capital expenditure, and nonoperational capital expenditure. These cash flows were discounted at 7% (the MMC's assumption about the cost of capital), which came to £457.9 million. Then the MMC added the present value of the opening less closing asset values of the distribution business, which represented another £128.2 million, giving a total of £586.1 million.

Asset values were calculated by taking an opening balance in 1990–1991 and rolling this forward by adding net distribution network capital expenditure. This was defined as network capital expenditure less depreciation. By the end of 1994–1995, this gave a total of £563 million and £610 million by the end of 1999–2000. The latter figure had a present value in 1995–1996 of £434.8 million.

The opening balance of £523.4 million in 1990–1991 was consistent with the figure used by the government in setting the original price control and the initial market value of HE. Table 2 shows the roll forward of the opening balance to £563 million at the start of the price control period in 1995–1996.

The total of £586.1 million in Table 1 represented the present value of the revenue that the MMC considered HE would need to raise to cover its allowable cash outflows and earn a 7% return on its asset value. The MMC calculated that the continuation of the existing price control would raise revenue with a present value of £462.1 million, which fell short of this amount. In the case of HE's distribution business, however, there was an additional source of revenue: the hydro benefit, which could be transferred from the generation business in accordance with HE's license. Taking this into account, the MMC decided that an appropriate relationship would be established and maintained if HE's price control required it to reduce prices by 0.3% in 1995–1996, followed by reductions of 2% per year for the next 4 years. Table 3 shows the MMC's projections of distribution business revenue. The present value of revenue and hydro

Table 1 MMC's calculation of HE's distribution business costs (1994/1995 prices)

	1995/1996	1996/1997	1997/1998	1998/1999	1999/2000	Total
Operating costs	60.7	59.5	58.3	57.1	56.0	
Network capital expenditure	43.5	43.2	43.8	44.1	44.6	
Non-operational expenditure	6.7	5.6	5.3	5.6	5.0	
Total	110.9	108.3	107.4	106.8	105.6	
PV of costs at 7%	107.2	97.8	90.7	84.3	77.9	457.9
PV of asset values a 7%	563.0				−434.8	128.2
						586.1

Table 2 MMC's calculation of HE's distribution asset base (1994/1995 prices)

	1990/1991	1991/1992	1992/1993	1993/1994	1994/1995
Opening value	523.4	534.6	534.4	536.1	545.1
Depreciation	27.2	27.9	28.7	29.7	31.0
Network capital expenditure	38.4	27.7	30.4	38.7	48.9
Closing value	534.6	534.4	536.1	545.1	563.0

benefit is £586.1 million, which is equal to the present value of costs and return on assets shown in Table 1.

The inflation index in the basic price restriction is generally one that is a good approximation of the previous year's inflation, reflects general price movements in the economy, is not focused on a particular segment of the economy, and is reliable and available in a timely manner. The regulator compares this price index with the average price change proposed by the company to determine whether the proposed price change is acceptable. The average price change is the weighted average change in prices, where a price's weight is the proportion of the company's revenue that the price generates.

Assume that a company has two services: service 1 and service 2. Service 1 provides 60% of the company's revenue, and service 2 provides 40%. The company proposes to increase the price of service 1 by 10% and the price of service 2 by 5%. The resulting average price change is: $(0.6\times10\%+0.4\times5\%)\times100=8\%$. If the basic restriction (inflation$-X$) is 8% or larger, the regulator approves the pricing proposal.

Extraordinary events may affect the utility disproportionately compared with the average firm in the economy. In these instances, regulators consider applying to the basic price cap formula an adjustment called an exogenous factor. Exogenous factors, also called Z-factors, reflect the effects of rare, one-time events whose occurrence and impacts are beyond the control of the regulated company and that affect the company differently from the average firm in the economy. An example might be a special tax placed on electric utilities. These exogenous factors increase or decrease the price index, depending on how the extraordinary event affected the utility.

SERVICE BASKETS

A service basket is a group of services placed under a common inflation$-X$ restriction. Services that the regulator wants to protect from price increases or decreases relative to certain other services are placed in a separate basket. If the regulator does not want the company to change urban prices relative to rural prices, for example, the regulator might place urban prices in one basket and rural prices in another.

The company is allowed to change the relative price levels of the services within a basket, subject to two possible restrictions. The first type of restriction is a limit on individual prices. Regulators may apply such a restriction by placing an absolute restriction on the price (e.g., the price per kWh for residential electricity cannot exceed 5 cents) or a percentage restriction (e.g., the price per kWh for residential electricity cannot increase more than 10% per year).

Regulators may also apply caps to subsets of services within a basket. The regulator may apply a restriction of inflation-5 to all services, for example, and a subrestriction of inflation-3 to residential services. In this case, the company's average overall price would need to decrease 5% in real terms, and residential prices would have to decrease 3% in real terms.

Table 3 MMC's projections of HE's distribution business revenue (1994/1995 prices)

	1995/1996	1996/1997	1997/1998	1998/1999	1999/2000	Total
Regulated revenue	105.2	104.6	103.8	102.9	102.1	
Unregulated revenue	5.5	5.3	5.1	5.0	4.8	
Hydro benefit	29.2	29.2	29.2	29.2	29.2	
Total	139.9	139.0	138.1	137.2	136.2	
PV of revenue at 7%	135.2	125.6	116.6	108.2	100.4	586.1

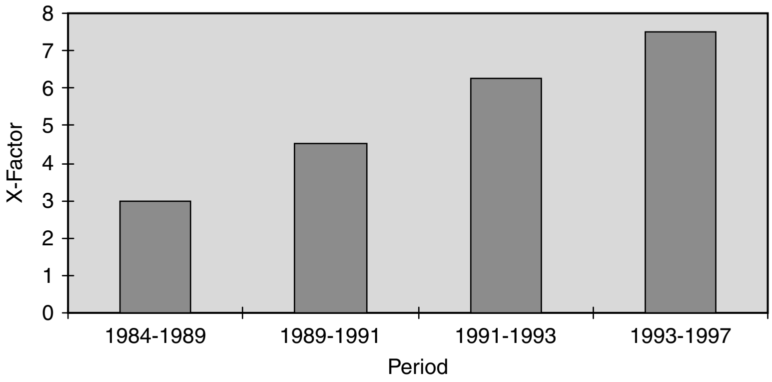


Fig. 1 Oftel's X-factors for BT.
Source: From Oxford University Press (see Ref. 11).

CASE STUDIES IN PRICE CAPS

Most applications of price cap regulation have been in telecommunications. Berg and Foreman[11] provide one of the earliest studies of the effects of price cap regulation, focusing on the U.K. regulation of British Telecom (BT), and the Federal Communications Commission's price cap regulation of AT&T and the Bell operating companies. The United Kingdom implemented price regulation for BT in 1984. There were four basic reasons why the United Kingdom adopted price regulation for BT:

- Price regulation would provide BT incentives to decrease costs.
- Because BT had been a government-owned service provider, information necessary for rate of return regulation was not available.
- The United Kingdom wanted to minimize the amount of adversarial litigation that had characterized U.S. rate of return regulation.
- The United Kingdom believed that regulation would serve primarily as a brief transitional mechanism to full competition.

The chart in Fig. 1 shows how the U.K. regulator changed the *X*-factor for BT over time. This was the general trend except for the 1997–2001 pricing decision. This growth in *X* may have related to the regulator's concomitant expansion of services covered under price caps (see Table 4), but it may also have reflected the regulator's increasing knowledge of how BT could improve its operating efficiency. Table 4 shows changes in services or elements subject to price control. Each price review has resulted in increasing numbers of services being subject to the price cap constraint. The chart in Fig. 2 shows the percentage of BT's turnover that is under price control for each period. This percentage grew from 48% during the first period to 71% during the 1993–1997 period.

Berg and Foreman[11] conducted their review using traditional rate evaluation criteria of simplicity and public acceptability, freedom from controversy, revenue sufficiency, revenue stability, price stability, fairness in apportionment of total costs, avoidance of undue rate discrimination, and encouragement of efficiency. They concluded the following:

- *Simplicity and public acceptability*. It is unlikely that price caps resulted in simplicity and administrative savings. Design of price caps required attention to service baskets and price bands, floors, and ceilings. The desire to increase public and other stakeholder acceptability created the need for additional control

Table 4 Changes in services subject to price cap

Period	Exchange line rentals	Domestic calls	Operator assisted calls	International calls	Connection charges
1984–1989	x	x			
1989–1991	x	x	x		
1991–1993	x	x	x	x	
1993–1997	x	x	x	x	x

Source: From Oxford University Press (see Ref. 11).

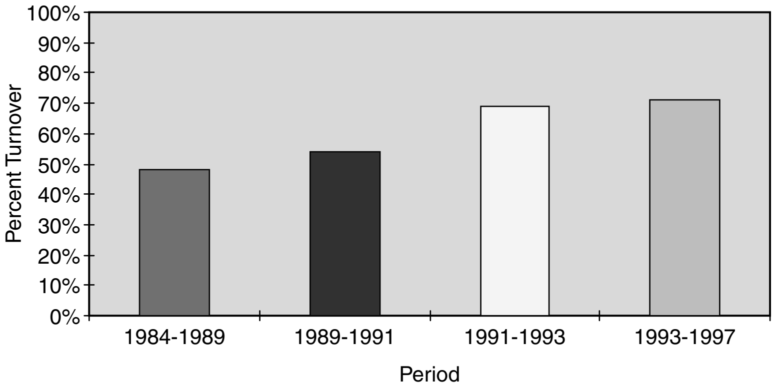


Fig. 2 Percent BT's turnover under price control.

features. Each feature has provided an opportunity for increased debate and litigation.

- *Freedom from controversy*. All the terms of price caps were controversial, including service quality, how to handle "excessive" returns, and public perceptions of the legitimacy of the regulation. Earnings-sharing assessments in the United States were sensitive to the same arbitrary cost allocations as rate of return regulation. Bell operating companies were given optional regulatory contracts and generally chose the lower productivity factors, even though these included higher earnings-sharing requirements.
- *Revenue sufficiency*. Competition complicated this objective. This would have been true regardless of the method of regulation. Weisman[12] concluded that regulators have less interest in revenue sufficiency when the price cap deal is struck.
- *Revenue stability*. This objective became one of net revenue stability (i.e., net income) under price cap regulation. As a result, using price caps increases the importance of making cross-subsidies explicit.
- *Price stability*. The price cap formula explicitly improves price predictability and stability relative to other prices in the economy by aligning price changes with changes in general inflation indices.
- *Fairness in apportionment of total costs*. Initial prices were part of a political compromise, so it was not immediately clear that price regulation results were different from rate of return regulation results with respect to this view of fairness.
- *Avoidance of undue rate discrimination*. Price caps use ceilings and floors to contain price discrimination. The U.K. regulator made three general changes to the price cap regime over time relative to this issue: (1) increased the *X*-factor, perhaps in response to high earnings by BT; (2) added special constraints to some prices, such as residential exchange line rental; and (3) added additional services and baskets (such as the median residential bill) over time.
- *Encouragement of efficiency*. British Telecom was allowed significant opportunity for rate rebalancing. Attenborough[13] found a total welfare gain of £2 billion per year in 1990–1991 prices, and 30% of this gain was from more efficient rate design. Price regulation allowed companies to improve economic efficiency by aligning prices with marginal costs, but competitive pressure, political constraints, and non-efficiency-related regulatory objectives may prevent this from happening in other situations.

CONCLUSION

Incentives and opportunities to improve efficiency are generally greater under price cap regulation than under rate of return regulation. This does not mean, however, that price cap regulation is the right form of regulation in all situations. Compared with rate of return regulation, price cap regulation decreases regulators' concern for revenue adequacy because they have less direct control over revenue. Also, regulators may come under pressure from consumer groups to break their commitment to allow higher earnings if the regulated company improves efficiency: consumers may view the higher profits as evidence that the regulator is not tough enough on the utility or is not knowledgeable. This challenge to regulatory legitimacy has led some regulators to roll back profits that they once said companies could keep.

When choosing a regulatory scheme, regulators should weigh these potential problems and benefits of price cap regulation against the corresponding costs and benefits of rate of return regulation. They may find that neither form of regulation is adequate by itself, and may adopt a hybrid system that applies different aspects of different forms of regulation to craft a regulatory scheme that makes sense for the regulator's institutional, political, and economic situation.[6]

REFERENCES

1. Lewis, T.R.; Garmon, C. *Fundamentals of Incentive Regulation*, 12th PURC/World Bank International Training Program on Utility Regulation and Strategy, Gainesville, FL, June 10–21, 2002.
2. Berg, S.V. *Introduction to the Fundamentals of Incentive Regulation*, 12th PURC/World Bank International Training Program on Utility Regulation and Strategy, Gainesville, FL, June 10–21, 2002.
3. Baldwin, R.; Cave, M. *Understanding Regulation: Theory, Strategy, and Practice*; Oxford University Press: Oxford, U.K., 1999.
4. Newbery, D.M. *Privatization, Restructuring, and Regulation of Network Utilities*; MIT Press: Cambridge, MA, 1999.
5. Lee, H. Price Cap: The U.K.'s Efforts to Regulate Regional Distribution Companies, Kennedy School of Government Case Program CR14-01-1619.0, Harvard University: Cambridge, MA, 2001.
6. Sappington, D.E.M. Price regulation. In *Handbook of Telecommunications Economics*, Cave, M.E. et al. Eds., North-Holland, Amsterdam: 2002; Vol. 1, 227–293.
7. Bernstein, J.I.; Sappington, D.E.M. How to determine the *X* in RPI-X regulation: A user's guide. Telecommunications Policy **2000**, *24* (1), 63–68.
8. Green, R.; Pardina, M.R. *Resetting Price Controls for Privatized Utilities: A Manual for Regulators*; World Bank: Washington, DC, 1999.
9. Bonbright, J.C.; Danielsen, A.L.; Kamerschen, D.R. *Principles of Public Utility Rates*; Public Utilities Reports, Inc.: Arlington, VA, 1988.
10. Office of Electricity Regulator. Transmission Price Control Review of the National Grid Company. Appendix D. Available at: http://cbdd.wsu.edu/kewlcontent/cdoutput/TR506/page29.htm (accessed on August 24, 2005).
11. Berg, S.V.; Foreman, R.D. Price cap policies in the transition from monopoly to competitive markets. Ind. Corp. Change **1995**, *4* (4), 671–681.
12. Weisman, D.L. Why less may be more under price-cap regulation. J. Regul. Econ. **1994**, *6*, 339–361.
13. Attenborough, N.; Foster, R.; Sandbach, J. *Economic Effects of Telephony Price Changes in the U.K.*, NERA Topics Paper No. 8; NERA Economic Consulting: London, U.K., 1992.

Pum–Ren

Regulation: Rate of Return

Mark A. Jamison
Public Utility Research Center, University of Florida, Gainesville, Florida, U.S.A.

Abstract
Rate of return regulation adjusts overall price levels according to the operator's accounting costs and cost of capital. In most cases, the regulator reviews the operator's overall price level in response to a claim by the operator that the rate of return that it is receiving is less than its cost of capital, or in response to a suspicion of the regulator or claim by a consumer group that the actual rate of return is greater than the cost of capital. Critical issues for the regulator include how to value the base, whether to add investments to the rate base as they are made or when the facilities go into service, the amount of depreciation, and whether expenditures have been prudently made and whether they relate to items that are used and useful for providing the utility service.

INTRODUCTION

This article describes rate of return regulation, which regulators often use to determine fair and reasonable prices for electricity sold by utility companies. Prices under rate of return regulation are considered fair and reasonable because they give the company an opportunity to recover the costs it has appropriately incurred in providing electricity service, and customers are protected from paying prices that would provide the company monopoly profits.[1] Rate of return regulation is sometimes criticized for not providing companies incentives to operate efficiently.

In performing this form of regulation, the regulator determines the appropriate amount for the company's rate base, cost of capital, operating expenses, and depreciation (Rate base is the gross value of the company's assets, minus accumulated depreciation. Cost of capital is also called the allowed rate of return and is the interest that the company pays on its debt, plus the return it must provide to shareholders to ensure that they continue to invest in the company). Based on these amounts, the regulator determines the amount of revenue the company needs to cover its operating expenses, depreciation, and cost of capital.

The emphasis on cost recovery in rate of return regulation is the source of the concern that companies may not operate efficiently.[2] If the regulator allows a rate of return that is higher than what the company actually needs to ensure that shareholders continue to provide capital for investment, the company could increase its returns to shareholders by making unnecessary investments (if the regulator does not catch the company doing so). This is called the Averch–Johnson effect.[3] Rate of return regulation, however, is also generally viewed as having the advantage of restricting opportunities for regulators to lower companies' prices arbitrarily.

Keywords: Rate of return; Asset; Rate base; Earnings; Cost; Revenue; Prudent; Used and useful.

BASICS FOR ASSESSING RATE OF RETURN

Assessing the company's rate of return involves evaluating the effects of price levels on earnings so that investors have an opportunity to receive a fair rate of return on their investments. There are five traditional criteria for determining whether a rate of return is appropriate.[1]

The first criterion is whether the rate of return is adequate to attract capital. One of the primary goals of utility regulation is to ensure that a sufficient level of service is available for customers. Service cannot be provided on an ongoing basis without continual reinvestment; therefore, capital attraction is a primary criterion for evaluating the rate of return.

A second criterion is the implementation of efficient management practices.

The third criterion measures efficient consumer-rationing of services. To encourage efficient consumption of services, prices should reflect marginal costs. In some situations, efficient prices may be different from those that attract capital, so there will be a need to balance these two criteria.

A fourth criterion is rate stability and predictability, which assists customers who value being able to plan their utility expenses.

The last criterion is fairness to investors. This may sound redundant with the capital-attraction criterion, but it is different. Consider a situation in which asset prices have been declining. A regulator may be tempted to adopt prices that are sufficient to attract investment for the new lower-priced assets but that are insufficient to recover the past costs of the historical assets. Such a decision would not satisfy the fairness criterion.

HOW RATE OF RETURN REGULATION WORKS

Rate of Return Regulation Basic Formula

Rate of return regulation combines a company's costs and allowed rate of return to develop a revenue requirement.

This revenue requirement then becomes the target revenue for setting prices. The basic formula for determining a revenue requirement is

$$R \equiv B \bullet r + E + d + T$$

where: *R*, revenue requirement; *B*, rate base, which is the amount of capital or assets the utility dedicates to providing its regulated services; *r*, allowed rate of return, which is the cost the utility incurs to finance its rate base, including both debt and equity; *E*, operating expenses, which are the costs of items such as supplies, labor (not used for plant construction), and items for resale that are consumed by the business in a short period (less than one year); *d*, annual depreciation expense, which is the annual accounting charge for wear, tear, and obsolescence of plant; *T*, all taxes not counted as operating expenses and not directly charged to customers.

Assume that the regulator determines that the company has a net asset base of $30 million, an after-tax cost of capital of 12%, a tax rate of 25%, operating expenses of $1 million, and depreciation expenses of $1,50,000. Further assume that the cost of capital is comprised of 50% debt and 50% equity; the cost of debt is 10%, and the cost of equity is 14%. (See the section on estimating the cost of capital later in this article.) To calculate the revenue requirement, we first need to determine the after-tax profit that the company would be allowed to receive when the new prices are in place:

Rate base (B)	$30,000,000
Cost of equity	× 14%
	$4,200,000
Percent of B financed by equity	× 50%
After-tax profit	$2,100,000

The company would pay a 25% tax on this profit—namely,

After-tax profit	$2,100,000
1–Income tax rate	÷ 75%
Pre-tax profit	$2,800,000
After-tax profit	− $2,100,000
Income taxes (T)	$700,000

We can now calculate the revenue requirement as:

Rate base (B)	$30,000,000
Allowed rate of return (r)	× 12%
Allowed after tax return (B × r)	$3,600,000
Expenses (E)	1,000,000
Depreciation expenses (d)	1,500,000
Taxes (T)	700,000
Revenue requirement (R)	$6,800,000

Advantages and Disadvantages of Rate of Return Regulation

There are several advantages to using rate of return regulation. The first is that it is sustainable if there is no competition, because prices can be adjusted to the company's changing conditions. It can also provide comfort to investors because rate of return regulation constrains the regulator's discretion in setting prices. This lowers investor risk, which lowers the cost of capital. Furthermore, company profits can be kept within acceptable levels from the perspectives of both investors and customers. Unless the regulator chronically underestimates the cost of capital (and courts do not reverse the regulator in this regard), investors can be confident that they have a fair opportunity to receive the profits they expect and, thus, are willing to make investments. Customers can observe that the regulator is limiting company profits to the cost of capital.

There are also several disadvantages to using rate of return regulation. First, it provides only weak incentives for companies to operate efficiently.[2] This weakness takes two forms. The first form is the Averch–Johnson effect. The second is that managers have less incentive to operate efficiently because regulators are unable to observe perfectly managers' efforts toward efficiency or the managers' innate abilities to be efficient. Another disadvantage is that rate cases have to occur frequently during times of high inflation (in the absence of a periodic adjustment for inflation between rate cases), and rate cases may be costly to perform. Last, rate of return regulation provides a mechanism for companies to shift costs from competitive markets to noncompetitive markets.

REVENUE IMPUTATION

In some situations, the company may receive revenue that should be attributed to the regulated operations but is recorded in the company's nonregulated operations accounts. In such cases, regulators can impute this revenue to the regulated operations. In the United Kingdom, for example, an electric distribution company receiving financial benefits from the government because the company used hydro power did not reflect this benefit in its regulated accounting books. Consequently, the regulator adjusted the revenue requirement to reflect this revenue.

The calculation for revenue imputation is fairly simple. First, the regulator calculates the revenue requirement. Then the regulator subtracts from the revenue requirement the revenue imputation amount. The adjusted revenue requirement is the amount of revenue the company is allowed to recover through prices charged for regulated services. Consider our previous example in which the revenue requirement was $5.275 million. If the regulator determined that there should be $2,00,000 in imputed revenue, the company would be required to charge prices designed to provide $5.075 million in revenue.

HOW RATE BASE IS DETERMINED

The Objective

When determining rate base, the objective is to identify the amount of capital the company uses and needs to use to provide regulated services. This capital includes the plant or facilities in service that the regulator determines to be prudent, as well as the working capital. The basic decisions include determining how to value the plant that is in service, for what time period the rate base is measured, and what plant is included. Each of these decisions is discussed below.

Methods for Valuing Plant in Service

There are three basic methods for valuing the plant that the company uses to provide its services.[1,4] The first of these is called fair value or economic valuation. There are two methods for determining fair value. The first is a nonmarket approach that bases fair value on the company's financial data, such as the discounted value of its cash flow. The second method uses the market valuation of the company. The fair-value approach can be problematic because it involves circular reasoning—namely, that the profitability of the company affects the asset value, which in turn affects profitability.

The most common method for valuing utility plant in the United States is the original-cost approach, or historical approach. In this approach, assets are valued at what the company originally paid to place it in service. If an electric company spent $5,00,000 for distribution wire and poles in 1999, and spent an additional $1,50,000 to construct the distribution facilities, that portion of the distribution network would be valued at $6,50,000, less depreciation, until it was retired from service.

The original-cost approach has the advantage of being objective because the values are tied to the financial records of the company, and provide a continual matching between the money the shareholders provide for investment and the cash flow that is provided back to investors. Disadvantages include difficulty of implementation where accounting and property records are poor, understating the economic value of assets during times of inflation, and providing misleading economic signals to markets as to the real economic costs of the electricity service.

The final approach to valuing assets is the replacement-cost approach, also called current-cost accounting, in which assets are valued based on what it would cost to replace them today. Under one version of this approach, the physical facilities in place are repriced at today's cost. If prices for wires, poles, and labor for the electric company in the previous example were increasing at a rate of 10% per year, the assets would be valued at $7,86,500, less depreciation, after two years. Replacement costs are generally determined by applying an inflation factor. They may also be valued by finding replacement prices in the marketplace. Because finding such prices is difficult, however, the inflation method is the most commonly used approach. Under an alternative version, sometimes called the virtual-company approach, the system is redesigned on paper to incorporate new technology not available when the original investment was made, and the components of this new system are valued at current prices.

Replacement-cost valuation has the advantage of being able to overcome some deficiencies of poor accounting records, although previous values are needed for applying the inflation approach. It also values assets near their economic value, which sends efficient price signals to customers and suppliers. The disadvantages include its subjectivity because of the difficulty of obtaining objective prices or inflation indices for old equipment, requiring exact inventories if current prices are used to estimate replacement costs, and returning to investors an amount of cash that is different from what they provided to the company. It is important to note that if rate base is defined in terms of current value, the cost of capital used to compute the revenue requirement should be adjusted to exclude inflation.

Some regulators use a combination of the historical and current-cost approaches. These regulators use historical costs for determining the total amount of revenue that the company is allowed to collect and current costs for designing prices. This combination gives the best of both worlds: investors receive back from customers exactly the cash they provided to the company, plus a return on their investment, and prices can send efficient economic signals to customers and managers.

Choosing a Test Period

A test period, which can be either a historical period or a future period, is the period of time chosen for quantifying the amount of plant that is being used to provide the utility service, expenses that are being incurred, and billing units that are being used to develop prices that will produce the allowed revenue requirement.[4] It is normally at least one year and may be several years. In choosing a test period, regulators generally attempt to choose a period that is representative of the time over which the prices will actually be charged. They also seek a period sufficiently long that it represents normal operations. In cases where the test period contains some unusual activities or events, such as an extraordinary amount of maintenance, regulators will generally normalize the costs or revenue that they believe may have been affected.

With historical periods, plant in service is measured for a recent period that is believed to be representative of the company's typical operations, for which the necessary accounting records are available, and for which all major adjustments to the accounting records have been concluded. This has an advantage of being objective and transparent, but the period for estimating plant in service is different from the period for which prices will actually be in effect.

When using a future period for the test period, plant in service is estimated based on projected changes. This has the advantage of being able to match the period for which plant in service is estimated with the period for which prices will actually be in effect. It has the disadvantage, however, of being subjective and may provide a greater incentive for companies to operate to meet the regulator's expectation rather than to be efficient. Alternatively, the company may project major additions to plant and then fail to make the investment and, in effect, receive a return and depreciation element on nonexistent plant.

Determining the Amount of Plant in Service During the Test Period

When the test period is chosen, the amount of plant in service may be valued in one of three ways: (1) average monthly balances, (2) end-of-period balance, or (3) average beginning-of-year and end-of-year balances. Average monthly balance simply estimates the arithmetic average of plant in service at the end (or beginning) of each month. End-of-period balance uses plant in service at the end of the year. Average beginning-of-year and end-of-year balances simply estimate the arithmetic average of the first month's and last month's plant in service.

Prudence Concept and Used and Useful Concept

In some jurisdictions, a utility plant must be considered both prudent and used and useful before being allowed into the rate base.[1,4] Prudent means the investment is reasonable based on cost-minimizing criteria. There are two perspectives. In one view, the investment is considered to be prudent if it was prudent at the time the decision was made. This requires accurately assessing what information management had available and used to make its decision. In the second perspective, the investment is prudent if management acted to minimize costs by fully considering changing conditions that would affect the investment. This requires assessing what management should have known and should have considered in making its decision.

Used and useful means that the plant is actually being used to provide service and that it is contributing to the provision of the service. If a company has excessive numbers of distribution lines carrying electricity to a neighborhood, for example, the regulator might not include some of the investment in the rate base, because even though all the lines are used, many are not needed, so they are not really useful.

Adjustments for Construction

Companies generally construct facilities over time. This means that investments are made prior to the plant's actually being used and useful. There are two ways to reflect this in the rate base.[1] One method, called construction work in progress (CWIP), includes the investment in the rate base as the investment is made. This is problematic because it violates the used-and-useful standard and causes current customers to pay for the plant that will be used by future customers. On the other hand, it provides cash flow for the construction project. The second method capitalizes the financing of construction projects and is called allowance for funds used during construction (AFUDC). AFUDC adds the cost of money used to finance the project during construction to the rate base once the plant is used and useful. Then the AFUDC is depreciated along with the plant. AFUDC does not provide cash flow to fund construction. The cash flow comes later and in some instances creates a cash surplus. This cash is returned to shareholders, held for future use, or invested outside the utility business. AFUDC makes current services more affordable to customers in the short run by capitalizing outlays and deferring returns until construction is complete. AFUDC can create "rate shock," however, when the full cost of the new plant plus accumulated financing costs enters rate base in one year.

Working Capital

Working capital is the average amount of capital, in excess of net plant, that is necessary for business operations.[4] Examples include inventories, petty cash, prepayments, minimum bank balances, and cash working capital. Cash working capital is the average amount of money that investors supply to bridge the gap between the time when expenses are incurred and the time when revenue is received. In some cases, customers prepay for services. This prepayment can be shown as an offset to cash working capital. When investors have provided working capital, it should be included in the rate base.

Accumulated Depreciation

Rate base excludes accumulated depreciation. In other words, the B in the formula includes the value of the plant less the amount by which it has been depreciated.

COST OF CAPITAL

The return the company is allowed to receive on its rate base is called the allowed rate of return or the cost of capital, and includes both the cost of debt that the company uses to finance its rate base and the cost of equity. The cost of debt is simply the weighted average of the interest rates that the company pays on its long-term corporate bonds. The cost of equity is the returns that shareholders need to ensure that they continue to finance the company. Regulators combine the cost of debt and cost of equity to form what is called the weighted average cost of capital (WACC).[1]

There are several ways to estimate the cost of equity, but the most popular is the capital asset pricing model (CAPM). Capital asset pricing model includes two basic components: the risk-free cost of capital and the risk premium. The risk-free cost of capital is generally considered to be the interest that the U.S. government pays on long-term bonds. The repayment of these bonds is generally considered to be secure, so the interest rate reflects only investors' time value of money. The risk premium is the amount of return that investors require because the actual earnings of the company are uncertain. This risk premium for a company is estimated by analyzing the degree to which the variation in the return on the stock follows the variation in the averaged returns on all the stocks in the market. When the costs of debt and equity are determined, regulators combine them into the WACC, using the company's capital structure as the weights.

OPERATING EXPENSES

Operating expenses[4] include costs of items such as supplies, labor (not used for plant construction), and items for resale that are consumed by the business in a short period of time (e.g., less than one year). Standards for accepting expenses include arm's-length bargaining, whether the expense is a legitimate expense for providing the utility services, whether the utility company has been inefficient or imprudent in incurring the expenses, and whether the expenses are representative. Arm's-length bargaining means that the company management, when deciding whether to incur an operating expense, has looked out for the financial interests of the company as though the company were in a competitive marketplace and had no financial interest in the expense payee, except as a purchaser of the payee's service or product.

Expenses are considered to be representative if they are being incurred at normal levels. Exceptional expenses may be disallowed if they do not represent the normal operations of the company. This applies primarily if the prices based on this revenue requirement will apply to multiple years. If this is the case, including expenses that are rarely incurred would cause these expenses to be recovered several times. Sometimes these expenses are normalized—that is, spread over multiple years. Also, expenses may be adjusted for known changes, such as pending wage increases or imminent decreases in numbers of employees. Expenses included in the revenue requirement are generally referred to as being "above the line." Expenses disallowed by the regulator are generally referred to as being "below the line."

DEPRECIATION

Depreciation is generally viewed as an annual accounting charge for wear, tear, and obsolescence. In regulation, depreciation is viewed as capital recovery—that is, the spreading of the plant investment over time to be recovered in revenue requirement.[4]

The determinants of depreciation are the useful life of the plant, salvage value, and depreciation method.[4] The useful life of the plant refers to the number of years over which the assets are depreciated. There are several ways to determine useful life. For many years, the physical life of the plant was used; this was usually longer than the economic viability of the plant, however. During times of technical and economic stability, actual experience with the length of time this type of plant is in use is appropriate for determining useful life. In most situations, though, historical experience does not match present or future needs because the stability preconditions are not met. Recently, the economic life of the plant, which is the projected length of time in which it will be economical to use the plant rather than replace it, has come to be favored. This approach has the down side of requiring accurate operating forecasts.

Salvage value refers to the market value of the plant at the time it will be removed from service, minus any removal and decommissioning costs. If this is a positive number, which means that the company can obtain some positive economic value for the asset when it is no longer used or useful, the salvage value is subtracted from the cost of the asset before depreciation is calculated. Assume that a building costs $20 million to build and that the company can sell it for scrap for $1 million after its useful life. The net cost to the shareholders for providing capital for the building is $19 million ($20 million minus $1 million); this is the amount that is depreciated. If, on the other hand, the building has contained hazardous waste, and it costs $5 million to clean up the site after the building is no longer in use and the materials are worthless, the amount to be depreciated is $25 million ($20 million plus $5 million).

Several methods are available for spreading the cost of the asset over time, but the most popular method is straight-line depreciation. With straight-line depreciation, the cost of the asset is spread uniformly over its useful life. Assume that the building will be used to store hazardous waste and will be used for 25 years. The annual depreciation using the straight-line method would be $1 million per year ($25 million divided by 25 years).

Straight-line depreciation is often criticized for not reflecting the rate at which the plant actually decreases in value. Generally, plant value decreases rapidly its first few years of service and more slowly in later years. Because straight-line depreciation assumes a constant rate of decrease, it understates actual depreciation in early years and overstates actual depreciation in later years. Slow depreciation rates create problems for companies whose markets are transitioning to competition or are experiencing increasing rates of technological change. If the regulatory depreciation rates are slower than economic depreciation, the company's book value of its plant may be greater than the economic value of its plant.

TAXES

Governments generally require utility companies to pay taxes, including income or profit taxes, franchise fees, property taxes, and excise taxes. Some of these taxes are passed directly to customers (it is common to see franchise fees and excise taxes listed on utility bills). Income taxes, however, are not passed on directly to customers, so it is necessary to include these taxes in the calculation of the revenue requirement. This is done by tax-effecting the revenue requirement. Consider a situation in which a regulator determines that a company's current revenue is $1 million less than the amount of revenue that would be needed to provide an adequate after-tax rate of return. Assume that the income tax rate is 20%. This means that the regulator would need to allow the operator prices that would increase before-tax revenue by $1.25 million ($1 million divided by 0.8, which is one minus the income tax rate).

The income taxes paid by the company may not match the income taxes that should be included in the rate base. This is because the timing in the accounting methods for tax purposes may be different from those used for regulatory purposes. Some countries may allow accelerated depreciation for tax purposes, for example, to encourage business investment. The regulator may prefer to use straight-line depreciation, however. The effect of using straight-line depreciation is that customers are in effect prepaying taxes and providing capital to the utility (one regulatory treatment, called "flow-through," eliminates this effect and removes the benefit to the utility of the favorable tax policy). This customer-provided capital typically receives one of two treatments in revenue requirements. Under the first treatment, the capital is deducted from the rate base. Under the second treatment, the capital is treated as cost-free capital.[5]

CONCLUSION

This chapter describes how regulators use rate of return regulation to control power companies' overall rate levels. Setting the overall rate level is just the first step in a two-step process for setting prices. The second step is rate design, which refers to the price structure or relationship among the individual prices.

In practice, regulators combine elements of rate of return regulation with other regulatory tools, such as benchmarking and price or revenue caps. These combinations allow regulators to customize the amounts of certainty and efficiency incentives that they believe are appropriate for their situation.

REFERENCES

1. Bonbright, J.C.; Danielsen, A.L.; Kamerschen, D.R. *Principles of Public Utility Rates*; Public Utilities Reports, Inc.: Arlington, VA, 1988.
2. Sappington, D.E.M.; Weisman, D.L. *Designing Incentive Regulation for the Telecommunications Industry*; MIT Press: Cambridge, MA, 1996.
3. Averch, H.; Johnson, L.L. Behavior of the firm under regulatory constraint. Am. Econ. Rev. **1962**, *52* (4), 1052–1069.
4. Phillips, C.F. *The Regulation of Public Utilities: Theory and Practice*; Public Utilities Reports, Inc.: Arlington, VA, 1993.
5. National Association of Regulatory Commissioners *Utility Regulatory Policy in the United States and Canada: Compilation 1995–1996*; The National Association of Regulatory Utility Commissioners: Washington, DC, 1996.

Renewable and Decentralized Energy Options: State Promotion in the U.S.

Steven Ferrey
Suffolk University Law School, Boston, Massachusetts, U.S.A.

Abstract
Renewable energy has emerged from the shadows of a novelty electric energy technology. In an era of fossil fuel volatility, renewable power which is immune from fossil fuel availability, pricing and delivery, makes power systems more resilient. Along with cogeneration, renewable energy can increase electric power efficiency, and renewable power can foster environmental benefits. Twenty-four of the 50 states promote these technologies with system benefit charges that underwrite renewable energy trust funds, and/or with renewable portfolio standards. These programs either subsidize or require a certain fraction of renewable electric power in the states' generation mix. Each of these states defines eligible renewable power differently, although all make wind power and solar power eligible. With implementation of these programs, the states have taken the lead on promotion of renewable energy.

MINIMIZING RISKS OF ENERGY SUPPLY DISRUPTION

The electric energy crisis in California during 2000–2001 demonstrated better than any study the unique and critical role of electricity in American society. In the course of a few months, the price of electricity quadrupled, skyrocketing the cost of living, and causing business dislocation. Nonetheless, blackouts rolled across the Golden State.

A shortage of electricity has dire social and political consequences—a blackout has been equated to a natural disaster. Allowing rolling blackouts is a tremendously inefficient way to balance supply and demand differences. Not every consumer attaches the same value to electricity at a given hour. For some industries, even a short blackout can ruin millions of dollars of production; for others, it is a minor inconvenience.

Since the attacks of September 11, 2001, the vulnerability of electric systems to systematic planned attacks has come under more analysis. The United States contains one quarter of all the electric generation capacity in the world. Because electricity cannot be easily stored or rerouted, supply must instantaneously match demand. System interruptions in the United States usually result from transmission and distribution difficulties within one-half mile of a customer. Where an electric system is centralized and integrated, a disruption from attack at a given point can temporarily destroy large parts of the integrated network.

Analysts argue that a distributed energy system, including increased use of cogeneration, is much less subject to disruption—whether it is from weather, terrorism, or other factors—than the centralized generation and distribution system employed in the United States. A distributed generation system typically is an electric generator placed on the consumer's side of the electric meter. It may be owned by the customer, the utility in rare instances, or a third party. By being so placed, it can either supply the host consumer, feed some or all power to the grid, or be wired to supply power to the host and selected abutters on dedicated distribution lines. When coupled with use of the thermal by-product of the generation process, distributed generation is known as cogeneration. A cogeneration system produces both electric energy and useful thermal energy. It thus uses some of the approximately 45%–70% of that portion of the output of a conventional electric power plant that is thermal energy and wasted as thermal pollution to the environment. Cogeneration is defined by federal law as a system that utilizes a minimum of 5% of the energy output as useful thermal energy while taking the remainder of the energy output as electric power. The robustness of a distributed, on-site, cogeneration-based system, likely fueled by natural gas, results from

- Reliance on a larger number of small generators; no one of which is critical to huge amounts of supply.
- Less reliance on a vulnerable centralized transmission and distribution grid.
- Reliance on the movement of natural gas in a more protected underground fashion to the electric generation source near the load center; rather than reliance on above-ground, more vulnerable electric transmission infrastructure. Gas can be stored in pipelines,

Keywords: System benefit charges; Renewable trust fund; Renewable Portfolio Standards; Green electricity; Renewable; Decentralized; Cogeneration.

while electricity cannot be stored in transmission lines, especially where they are knocked out.

EFFICIENCY AND CONTROL

From an efficiency point of view, there are significant reasons to promote decentralized on-site electricity supply. Decentralized electric production can transform electric production efficiency from approximately 33% for central station conventional utility supply to something approaching 80% for decentralized cogeneration. These decentralized electric supply technologies, in addition to greater potential efficiency and in certain circumstances, environmental benefits, tend to encourage the deployment of renewable energy sources and applications.

Because renewable energy sources are not under the control of any nation or cartel but are distributed across the earth, they are not subject to embargo or manipulation. Because decentralized renewable energy sources are developed in relatively small modules, this promotes reliability and resiliency of the system. Because decentralized energy resources are built close to their points of use, they are not as dependent on long transmission and distribution networks and are less vulnerable to supply disruption from an overloaded system line, storm, or intentional disruption.

Unlike finite fossil fuels, solar energy represents a constantly replenished flow, rather than an existing stock that is diminished by its use. By contrast, burning a barrel of oil or a cubic meter of natural gas diminishes permanently that quantity of fossil fuels for the next day and for future generations. While many nations—particularly developing nations—have no significant reserves of oil, coal, or natural gas, every nation has solar energy in some form—sunlight, wind, or ocean wave power.

GREATER EFFICIENCY OF COGENRATION AND DISTRIBUTED ENERGY OPTIONS

Both conventional electric generation technologies and industrial process heat applications are inefficient. Conventional electric generating technologies typically exhaust as much as two-thirds of the heat energy produced to power electric generators. Many industries use process steam in applications that run below 400°F; however, the combustion of fossil fuels required to produce that heat results in temperatures of more than 3000°F, much of which is wasted. The next major leap in efficiency must come from recovering and reusing waste heat. Machines that recover all waste heat and produce electricity have the capability to achieve efficiencies from 50 to 90%, much better than the typical 30+% of the existing central station utility fossil steam system. Thus, cogeneration facilities operate at overall thermal efficiencies as great as 250%–300% higher than conventional electric generating technologies. The very best cogeneration technologies are more than twice as efficient as new coal-fired power plants. This increase in efficiency results in savings of up to 31% in the fuel input needed to generate a unit of usable energy output by various cogeneration technologies when compared to conventional electricity generation technologies.

Distributed generation and cogeneration systems, because they are smaller, tend to be less efficient at electric production than large central station power generation facilities. It has been estimated that distributed generation would be 23% more expensive to implement in Florida (where cooling requirements dominate) and 27% more expensive to implement in New York state (where heating requirements dominate) than a new centralized system.

However, this comparison looks only at electric production. If one assumes that waste heat from distributed cogeneration can be employed productively, the economics change: a distributed cogeneration model realizes cost savings of 30 and 21% in New York and Florida, respectively. Interestingly, as the cogeneration units get smaller, total system savings increase. This is due to the improved cost profiles of new gas turbines and internal combustion engines, and the ability of smaller units to meet variable loads more efficiently. Also, small units can be sited where waste heat can be most productively used. Thus, the cogeneration value of distributed generation turns the economics from negative to positive because of the greater overall efficiency of energy production and use.

The heat recovered from a total cogenerating energy system can be used for direct application heat, for industrial process heat, or for preheating the combustion air for a utility boiler. This means that more useful energy can be produced while generating fewer environmental pollutants and emissions. It also means that less transmission capability would be required if there is development of dispersed electric and total energy systems, located close to load centers. Not only will additional transmission capacity not be required in certain areas, but capacity on existing transmission grids will be less burdened. One way to view this phenomenon is that if natural gas cogeneration or total energy systems replace centrally dispatched electricity, energy will be moved more in its primary form by natural gas pipelines and less in its derived form as electricity.

ENVIRONMENTAL BENEFITS

Conventional production of electricity by electric utilities in the United States is responsible for substantial shares of criteria pollutant emissions, including 68% of SO_2

emissions, 33% of NO_x emissions, and 33% of CO_2 emissions. Environmental costs associated with power plants occur at each of three stages of the energy process:

- Front-end costs at the point of extraction and processing of energy sources. These include the costs of drilling, mining, or otherwise extracting raw fuel sources; the processing, enrichment or concentration on these fuel sources; the manufacture of equipment to effectively utilize these fuel sources; and transportation costs for fuel and equipment.
- Direct costs associated with the use of energy sources. These include the emission of a variety of pollutants, health impacts from these emissions, impacts on the natural environment of such emissions, and human occupational exposure or illness at the power plant work site. The primary effects on human populations are the increased risk of mortality and morbidity, including chronic illness and increased risk of chronic disease.
- Back-end residual costs. These include waste disposal costs for residual elements of fuel and the eventual costs of decommissioning energy producing facilities.

The primary impacts on human health from direct production of electric energy are from emissions of the criteria pollutants carbon dioxide, sulfur dioxide (SO_2), NO_x, ozone, and particulates; and from acid deposition. Carbon dioxide is caused principally by the burning of fossil fuels, and it is a principal greenhouse gas responsible for global warming. Sulfur exerts a significant impact on human health directly, it is a precursor of aerosols that result in acid deposition, and it is transformed into sulfates, which pose independent problems. NO_x is formed by the conversion of chemically bound nitrogen in the fuel or from thermal fixation of atmospheric nitrogen in the combustion air. Ozone causes damage to human health, agriculture, and plant life. Particulates include solid particles and liquid matter which range in size from 1 micron to more than 100 microns in diameter. They are responsible for major health impairment, impairment of visibility by causing haze, and the creation of sulfated from SO_2 emissions. Acid deposition causes damage to forests, wildlife, water quality, and aquatic species. Conventional power facilities exert environmental impacts on health and the environment in the form of water pollution and impairment of land uses. The choice of fuels as well as the technology used for converting those fuels to electricity has profound implications for attaining CO_2 reduction targets to limit possible effects of global warming.

Cogeneration facilities should cause fewer environmental impacts than equivalent megawatts of conventional power production because cogeneration facilities simultaneously produce electricity and thermal energy by the same process, thereby recapturing and utilizing energy that would otherwise be wasted. This substitution of an integrated cogeneration technology for conventional separate electricity and thermal energy production technologies should save 15%–25% of the energy input otherwise consumed by separate energy production configurations, according to government studies.

Typical air emissions of technologies, without added emission controls, are displayed in Table 1.

STATE OPTIONS FOR PROMOTION OF RENEWABLE TECHNOLOGIES

There are several recognized techniques capable of deployment to promote renewable energy and demand-side management (DSM) investments. Each attempts to require or subsidize certain preferred technologies that otherwise might be less demanded by the market.

System Benefits Charge/Renewable Trust Funds

The system benefits charge is a tax or surcharge mechanism for collecting funds from electric consumers,

Table 1 Air quality impacts of cogeneration

Technological characteristics	Direct physical effects	Impact on air quality: positive, negative or mixed
Increased efficiency	Reduction in total emissions per unit of energy produced	Positive
Smaller scale of QF deployment	Change in emissions levels	Mixed or negative
	Change in level of environmental control, usually less	Negative
	Lesser emissions stack height	Mixed or negative
Change in energy production technology	Change in emissions and type of pollutants	Mixed
Change of fuels	Change in emissions and type of pollutants	Mixed or positive
Change of location of electricity generation	Change in location of emissions, density, and distribution	Mixed

Table 2 Funding levels and program duration

State	Approximate annual funding (millions of dollars)	Per capita annual funding	Per MWh funding	Funding duration
California	$135	$4.0	$0.58	1998–2011
Connecticut	$15–$30	$4.4	$0.50	2000–indefinite
Delaware	$1 (maximum)	$1.3	$0.09	10/1999–indefinite
Illinois	$5	$0.4	$0.04	1998–2007
Massachusetts	$30–$20	$4.7	$0.59	1998–indefinite
Montana	$2	$2.2	$0.20	1999–7/2003
New Jersey	$30	$3.6	$0.43	2001–2008
New Mexico	$4	$2.2	$0.22	2007–indefinite
New York	$6–$14	$0.7	$0.11	7/1998–6/2006
Ohio	$15–$5 (portion of)	$1.3	$0.09	2001–2010
Oregon	$8.6	$2.5	$0.17	10/2001–9/2010
Pennsylvania	$10.8 (portion of)	$0.9	$0.08	1999–indefinite
Rhode Island	$2	$1.9	$0.28	1997–2002
Wisconsin	$1–$4.8	$0.9	$0.07	4/1999–indefinite

Source: From U.S. Department of Energy (see Ref. 1).

the proceeds of which may then support a range of activities. In order to support DSM or renewable resources, funds are collected through a nonbypassable system benefits charge to users of electric distribution services. The money raised from the system benefits charge is then used to “buy down” the cost of power produced from sustainable technologies on both the supply and demand side, so that they can compete with more conventional technologies. A system benefits charge will raise the following issues: the level of the charge, the allocation to classes of customers, the rate design, the programs to be implemented, and the ongoing process for oversight and management of the fund.

Between 1998 and 2012, approximately $3.5 billion will be collected by 14 states with existing renewable energy funds. More than half the amount collected, at least $135 million per year, comes from just California. The funding levels range from $0.07/MWh in Wisconsin up to almost $0.6/MWh in Massachusetts. Most only provide assistance to new projects, and not existing renewable projects.

The form of administration of renewable trust funds varies. Many states administer them through a state agency, while others use a quasi-public business development organization. Some funds are managed by independent third-party organizations, and some are managed by existing utilities, while two states allow large customers to self-direct the funds. For distribution, some states utilize an investment model, making loans and equity investments. Other states provide financial incentives for production or grants to stimulate supply-side development. Some other states use research and development grants, technical assistance, education, and demonstration projects.

As Table 2 indicates, the funding level is in the range of $175–$250 million annually for the cumulative impact of the fourteen state system benefit charge programs. While many of these programs are set up to run indefinitely, others have set life spans. The level of per capita funding ranges between $0.90 and $4.40 annually for renewable energy. Expressed another way, for each megawatt-hour sold in the state, the level of subsidy ranges from $0.07 to $0.59.

Renewable trust funds are likely to be less efficient than portfolio standards in promoting the burgeoning renewable power industry. Portfolio standards set a requirement and challenge market participants to satisfy it in the most efficient manner possible. By contrast, trust funds create a discretionary gift program. This process will cause renewable projects to conform themselves to funding criteria rather than to take the initiative to operate most efficiently. Political manipulation of trust fund cash flows also is possible.

Renewable Resource Portfolio Requirements

A resource portfolio requirement requires certain electricity sellers or buyers maintain a predetermined percentage of designated clean resources in their wholesale supply mix. A number of variations of resource portfolios are possible, including a renewable resource portfolio requirement, a DSM portfolio requirement, and a fossil plant efficiency portfolio requirement.

While Massachusetts and Connecticut are the first two states in the initial wave or retail deregulation to adopt both a system benefit charge to fund renewable technologies and a resource portfolio standard mandating renewable wholesale power sources, 22 states and the District of Columbia have adopted the renewable portfolio standard. The key to making the portfolio requirements work is to establish trading schemes for "portfolio obligations." A trading scheme would allow distribution companies, energy service companies (ESCOs), or generation companies that are particularly effective at developing low-cost DSM programs or renewable resources to sell portfolio obligations to other distribution and generation companies that are less effective in developing these resources. (ESCOs are companies that achieve conservation savings on customers' premises, and often split such savings with the customer or charge a fee for service.)

Portfolio standards are flexible, in that certain technologies can be included in the renewables definition or certain subgroups of technologies can be targeted for inclusion at distinct levels. The standard allows market competition to decide how best to achieve these standards. The standards become self-enforcing as a condition of retail sale licensure.

The advantage of a portfolio standard is that it does not subsidize any particular technology or locus of that technology. Resource portfolio requirements can be applied under any wholesale or retail competition, without placing any entities at a disadvantage.

A disadvantage to implementing portfolio resource requirements is that the appropriate portfolio target level must be decided on, rather than relying on general financial incentives towards continuous improvements. The primary disadvantage for DSM standards is the logistical challenge of monitoring tradable obligations for DSM, which requires continuing regulatory oversight for measuring and monitoring DSM savings. A renewable standard will also involve subsidiary issues, such as how to define renewable resources and whether standards should be based on available capacity or actual generation; though from an environmental perspective, basing them on actual kWh of generation makes the most logical sense.

Table 3 Portfolio standards and trust funds in deregulated states

State name	Renewable energy trust fund	Portfolio standards
Arizona		x
California	x	x
Colorado		x
Connecticut	x	x
Delaware	x	x
District of Columbia	x	x
Hawaii		x
Illinois	x	x
Iowa		x
Maine	x	x
Maryland		x
Massachusetts	x	x
Minnesota	x	x
Montana	x	x
Nevada		x
New Jersey	x	x
New Mexico		x
New York	x	x
Ohio	x	
Oregon	x	
Pennsylvania	x	x
Rhode Island	x	x
Texas		x
Vermont		x
Wisconsin	x	x

"Green" Electricity Pricing

Certain commodities with an "environmental" identification sell for more than generic commodities. This is true for organic foods, "pure" spring bottled water, and products made from recycled parts, for example. The average price of bottled water is 1000–4000 times the cost of tap water in the U.S. Bottled water often is tap water in a bottle. This indicates that certain brand identities or green products command a substantial premium in the market of commodities.

The operational premise of so-called "green pricing" is that there are certain customers who want to use electricity produced by "clean" or renewable technology and are willing to undertake the expense in order to procure it. Electricity sellers can then use funds raised from such environmentally oriented customers to acquire less-polluting resources that would not otherwise be developed because of market costs or market barriers. One advantage of green pricing is that providers can protect customers from fuel price fluctuations, as most renewable resource prices do not experience the volatility of oil and gas-fired electricity, which fluctuate with fossil fuel commodity prices.

Renewable power that qualifies as "green" energy can be marketed and sold in different modes:

- Actual kWh sales, in which electricity and its green attributes are bundled together as a single product in a single transaction; there, the purchaser switches from

conventional power to a bilateral contract with the green energy supplier.
- Sale only of the green attributes without purchasing the actual kilowatt hours; the retail customer does not need to switch from buying conventional power but only purchases the renewable certificate.

The drawback of green power marketing is that it relies on individual consumer decisions to create a public good. The environmental benefits of green power consumption are not internalized to the consumer who elects to pay a premium for green power, but rather are shared by all in the region. This allows free riders to benefit without having to pay or contribute. This raises equity and efficiency issues. Therefore, green power marketing suffers from individual consumer motivation and equity impediments that are not raised by a portfolio standard, which imposes requirements on energy producers rather than consumers.

In the deregulated Pennsylvania market, over 500,000 residential customers switched to competitive suppliers. Of those, 15%–20% chose green electricity products, paying a premium of 0.8–1.5 ¢/kWh over standard electricity pricing. In California, almost 100% of customers who switched retail suppliers prior to the California restructuring debacle consumed a green power

Table 4 "Renewable" resources as defined in state statutes

State	Solar	Wind	Fuel cell	Methane/ landfill	Biomass	Trash-to-energy
California	x	x		x	x	x
Connecticut	x	x	x	x	x	x
Illinois	x	x			x	x
Maine	x	x	x		x	x
Massachusetts	x	x	x	x	x	x
Nevada	x	x	x			
New Jersey	x	x	x	x	x	x
New Mexico	x	x	x	x	x	x
New York	x	x		x		x
Oregon	x	x		x		x
Pennsylvania	x	x		x	x	x
Rhode Island	x	x		x	x	x
Texas	x	x		x	x	x
Wisconsin	x	x	x		x	x

State	Hydro	Tidal	Geothermal	Photovoltaic	Dedicated crops
California	x		x	x	
Connecticut	x			x	
Illinois	x			x	x
Maine	x	x	x	x	
Massachusetts	x	x		x	x
Nevada			x	x	
New Jersey	x	x	x	x	
New Mexico	x	x	x	x	
New York	x	x	x	x	
Oregon	x	x	x	x	x
Pennsylvania	x		x	x	x
Rhode Island	x			x	
Texas	x	x	x	x	
Wisconsin	x	x	x		x

Note: "Photovoltaic" is likely included within "solar" in some states; "methane" and/or "trash-to-energy" may be included within a broad definition of "biomass."

product, paying a premium of 1.1–2.5 ¢/kWh over a standard electricity price. In California, this resulted because of a state renewable energy credit of 1.5 ¢/kWh to green power suppliers, allowing them to undercut the default service price. When the California market collapsed, almost all of these green consumers were returned to standard utility service. By 2002, "green-e" certified products were available in Connecticut, New Jersey, Pennsylvania, and Texas, offered by four suppliers. Where utilities sell green power products, participation rates are typically 3%–4%, with customers paying premiums ranging from 1 to 7.6 ¢/kWh.

While public opinion surveys reveal that 50%–95% of Americans say they are willing to pay more for renewable power, in fact, only approximately 1% of residential consumers who have an ability to select and purchase renewable energy power have done so. On the other hand, bottled water reached 8% market penetration, socially responsible investing a 13% market penetration, and recycling a 25% market penetration over an extended period time of 1–2 decades from the original product launch. Customer participation rates in green power programs, as of 2001, were less than 1% in more than half of programs available.

The ability of green pricing mechanisms to promote cleaner resources will depend on how willing customers are to pay for them, as well as the success of the electricity sellers' marketing and promotion campaigns. Any rate impacts associated with environmental improvement are assigned to those customers who choose to accept them. As a result, green pricing does not cause any market distortions and is, in theory, reflective of true customer preferences. Green pricing effectively assigns an important social policy decision to individual customers, who may act on their own interest rather than in society's best interest.

STATE PROGRAMS FOR RENEWABLE ENERGY

As of 2004, 22 states and the District of Columbia had enacted either a law or an administrative order to create open power markets, thus allowing consumers to choose their electricity supplier. States deregulating their retail electric sectors have implemented renewable portfolio standards and/or trust funds. Twenty four states have elected one or both of these options (see Table 3).

The renewable resource measures that states have incorporated into electricity restructuring and deregulation statutes vary. Some renewable energy measures create portfolio standards; others create trust funds to invest in the development and utilization of renewable resources. Some adopt both concurrently. How each defines an eligible renewable resource varies significantly.

Table 3 illustrates how states have deployed these two options. Each defines what an eligible renewable resource is differently. The diverse pattern of "renewable" resources included under state definitions is set forth in Table 4.

REFERENCES

1. Bolinger, M.; et al. Clean Energy Funds, LBNL-47705 (2001) at ix.

Renewable Energy

John O. Blackburn
Emeritus of Economics, Department of Economics, Duke University, Durham, North Carolina, U.S.A.

Abstract
Renewable energy sources in the United States include solar heat and electricity and indirect solar sources like wind, falling water, biomass, wave, ocean current, and ocean temperature differences, as well as geothermal and tidal possibilities. The first two sources, solar radiation and wind, are very large. Solar radiation exceeds total U.S. energy use by a factor of 440, while the wind energy potential is three times the annual electricity use. Hydroelectric, biomass, and geothermal sources are commercially developed, as well, and can supply half the nation's energy. The largest sources are also intermittent in nature, thus raising issues of complementarity and storage in high-renewable energy scenarios.

U.S. energy use, now in the 100-quad range, can be greatly reduced by efficiency gains. Energy end-uses in the forms of heat, liquid fuels, and electricity are examined. Reduced energy demands can be met largely or wholly via renewable sources, thus preserving relatively high living standards in a sustainable manner, given time, commitment, and further development of two or three key technologies.

INTRODUCTION

Renewable energy is a category of energy supply that is receiving increasing attention in the United States. The term refers, of course, to energy sources that can be fully utilized without diminishing future supplies. The obvious contrast is with fossil fuels which, once used, are no longer available.

Renewable energy is certainly nothing new. Mankind lived on it almost exclusively until the industrial revolution in the 1700s. Falling water and wind provided mechanical energy for tasks like pumping water or grinding grain, while ships were propelled by wind. Wood fires provided heat for dwellings or for tasks like smelting metals.

This article examines the various renewable energy resources in the United States and the status of conversion technologies. Because the largest two sources are also intermittent, integration into reliable supply systems is also examined. This turns out to be a less formidable task than it first appears to be.

Present energy end uses are examined in order to see how renewable sources might be deployed in a highly renewable energy future.

RENEWABLE ENERGY FORMS

Renewable energy comes in many forms, though most sources come directly or indirectly from solar radiation. Solar energy may be used directly as a source of heat as, for example, in heating water. It may also produce electricity, either directly in photovoltaic cells or by using concentrating collectors to produce steam for turbines. Wind energy, which is derived from solar energy, is now used almost exclusively to produce electricity, as is falling water. Winds over ocean waters produce waves, which are beginning to be harnessed as a source of electricity. Ocean currents or temperature differences between surface and deep waters also offer opportunities to extract useful energy.

Trees, crops, and other plant matter (referred to as "biomass" in the renewable energy literature) likewise represent stored solar energy. Half the world still uses wood as a fuel, while forest product industries in the industrial world use wood waste to produce process heat, steam, and electricity. Energy crops, trees, or even grasses could be grown. Fuel alcohols are now produced mostly from sugarcane and corn, and they are beginning to be produced from cheaper, more abundant cellulosic biomass materials, such as wheat straw or other waste products. Other biomass sources include animal manures and sewage, which can produce methane through anaerobic digestion. There are a few renewable sources of energy that do not derive from solar sources.

The moon's gravitational pull on the earth creates tidal movements of ocean waters. In some locations, the rise and fall of water levels is sufficiently great to generate electricity.

Heat generated from nuclear processes deep in the earth sometimes comes close enough to the surface to be captured as useful energy. When underground water is present as well, steam or hot water may be available for electricity generation.

The principal difference between age-old uses of renewable energy and the present possibilities is the

Keywords: Solar energy; Wind energy; Renewable resources; Biomass energy; Renewable energy future; Intermittent energy sources.

availability of modern technologies with conversion efficiencies much higher than those previously available. Most importantly, all forms of renewable energy can be converted into electricity—the most versatile form of energy for human purposes.

POTENTIAL RENEWABLE RESOURCES IN THE UNITED STATES

Solar radiation falling on the lower forty-eight states amounts to an estimated 44,000 quads (quadrillion Btu) annually — 440 times the annual energy use for the nation.[1] Few other resources are found in such abundance. Windpower, the next most abundant renewable energy source, has an estimated electricity-generating potential of 10,000 billion kilowatt-hours in the United States, almost three times present annual electricity consumption.[2] This estimate is for on-shore wind turbines only; the additional offshore potential is large, but has not been as carefully estimated. Moreover, recent studies find the U.S. potential to be underestimated in quantity and in the number of areas with economic wind-generating potential.[3]

Hydroelectric facilities already generate 7%–8% of the nation's electricity, with another 3%–4% that might still be developed.[4] Geothermal resources that are suitable for electricity generation are found only in the western third of the United States, but they amount to 25,000 megawatts—enough to provide 25%–30% of that area's present electricity use. Geothermal heat is available to more of the nation, with a much larger potential.[5] The wave-generating potential has not yet been fully assessed in the United States. It appears to be substantial along the Pacific coast and the North Atlantic coast. Tidal power possibilities in the United States are limited to Alaska, Maine, and Washington.

As for biomass sources, U.S. forests and farms now produce materials with an energy content of some 14 quads. With energy crops, this could increase fourfold, according to estimates of the National Renewable Energy Laboratory.[6] Other recent estimates give a near-term potential of 19–20 quads over and above materials taken for food, feed, and forest products, given relatively minor modifications in farm and forestry practices.[7]

There is clearly no shortage of renewable energy. It comes in many forms and, with respect to solar radiation and windpower, in great abundance.

In order to be useful in human activities, renewable energy must be captured and converted into forms which are consistent with energy demands. Time gaps between availability and need must also be bridged. For these tasks, equipment that is reliable, durable, and affordable must be available.

CONVERSION EFFICIENCIES

Conversion efficiencies for various forms of renewable energy vary considerably. The growth of trees and other plant life generally convert only 0.5%–1% of solar radiation into stored biomass energy, which can be used commercially. The fastest growing species in areas with long growing seasons (sugarcane, for example) convert 2%–3% of solar energy into recoverable energy.[8]

Other renewable technologies, fortunately, have much higher conversion ratios. Solar panels for heating water convert 40%–55% of the solar energy that falls on their surfaces to useful energy.[9] Crystalline silicon photovoltaic cells now on the market convert 11%–14% of solar radiation into electricity, lab models reach 33%, and nanotechnologies hold some promise of yet-higher figures.[10] It is these conversion efficiencies that facilitate the use of large quantities of renewable energy collected in a relatively small portion of a nation's total area.

THE STATE OF TECHNOLOGIES: COSTS

Technologies Now Fully Developed and Cost Competitive

Hydroelectricity has been competitive since the earliest days of grid electricity. Expansion in the United States is constrained by environmental considerations, though output can be expanded by repowering existing facilities or installing generators at small existing dams. Hydro capacity becomes extremely valuable in grids with high components of intermittent power from wind and solar sources because it enables them to operate over a considerable range without further backup.

Wind energy is competitive with conventional fuels in areas with good wind resources and its costs continue to fall.

Geothermal Electricity

This technology is now well established in suitable sites around the world. Initial capital costs are high, both for drilling into steam or hot water sources and for generating equipment. California has 2200 megawatts in operation, and new facilities are planned or under construction in several Western states.[11]

Biomass

About three quads are already used annually for energy in the United States. The forest products industries use 1.8 quads of wood and wood wastes for process heat and cogenerated electricity. Firewood heats three million homes and supplies some occasional heat to another 20 million. These uses are fully competitive with fossil fuels, as is methane collected from landfills.[12] Other biomass possibilities are listed in the following section.

Ocean Sources

A small number of tidal plants have been installed around the world, but none have been installed as of yet in the United States. Other technologies for using ocean energy, except for waves, are not under active development at this time.[13]

Technologies that are Well Developed but not Yet Fully Cost Competitive

Solar Water Heating

This is a relatively simple technology consisting basically of water pipes in a box covered with glass and painted black. Today's versions are, to be sure, much more sophisticated, with freeze-protection features to make the devices useable nationwide. Making costs competitive with present energy sources will be possible upon mass production and mass installation of the equipment, a feat long since accomplished in Israel.

Photovoltaic Electricity

Fifty years of research and development efforts have brought higher efficiencies and greatly reduced costs. Costs are still too high, by a factor of two or three, for photovoltaic electricity to be fully competitive on customer facilities in the United States. They are already competitive for users remote from a power grid or for utilities at peak summer hours.

Cost reductions in photovoltaic systems manufacturing are proceeding more slowly than anticipated in the 1980s and 1990s, while conventional generation, with which these systems compete, has shown declining inflation-adjusted cost, as well. Authoritative analyses in the 1980s projected cost reductions of photovoltaic modules and other system costs to about one dollar per peak watt (two dollars in 2005 prices), which implied Photovoltaic (PV) electricity costs of six to seven cents per kilowatt-hour.[14] Module prices have indeed declined to about $3.50 for volume purchases in 2005 (in current $).[15] Costs are not likely to fall significantly until some alternatives to the (now) standard silicon wafer are developed for volume production. Nonetheless, manufacturing costs have fallen below three dollars for modules.[16] With rapid market and manufacturing development, Japan can produce photovoltaic electricity at about 15 cents per kwh, competitive with (expensive) Japanese-delivered grid electricity.[17]

Photovoltaic equipment benefits from its potential location at the point of consumption, thus avoiding most of the costs of transmission and distribution. PV electricity, therefore, can be priced higher than electricity from central generating plants. Because there is already enough rooftop space to accommodate nearly any conceivable volume of photovoltaic power, site costs are avoided as well.

Other Biomass Sources

Subsidized fuel ethanol is made mostly from corn, but more abundant cellulose sources are beginning to be utilized. Scale economies and cost decreases with production experience should make this technology competitive in the next few years. Iogen, the Canadian producer of ethanol from wheat straw, estimates a price of $1.30 per gallon from a proposed plant in Idaho.[18] Economically feasible processes and equipment to produce methane from animal wastes are already available, but the technology is still not widely used.

Technologies for Which Further Development is Required

Solar Industrial Process Heat

Development of equipment for these tasks received some attention in the late 1970s, but was abandoned after that as fuel prices declined. Industrial process heat requirements and insolation overlap considerably in the United States, so their potential remains to be exploited.[19] Further development awaits higher fossil fuel costs and some renewed research attention.

Wave Energy

Several small demonstration facilities now operate. The first full-scale commercial generation facilities are about to be installed in Portugal by British and Portuguese firms using British designs.[20]

Ocean Currents and Thermal Differences between Surface and Deep Ocean Waters

There is very little activity at present.

In summary, hydroelectric, wind, and geothermal electricity are fully cost-competitive with fossil and nuclear-generated electricity in areas with good resources. These areas are, respectively, three-quarters of the states for wind, all but three or four states for hydroelectricity, and the western third of the nation for geothermal. Biomass-based heat and electricity are likewise well-established sources and they are available in every state. Widespread market penetration of solar hot water systems awaits economies of mass production and mass installation. Biomass-based liquid fuel production is set to expand rapidly. Sales of photovoltaic equipment are rising rapidly, but still with subsidies and incentives. Cost reduction proceeds apace at about 4%–5% per year, while several promising new and potentially much cheaper versions are moving into production.

Solar industrial process heat is receiving little attention, though it probably will as oil and gas prices trend upward in coming decades.

DEALING WITH INTERMITTENT SOURCES

Solar radiation and wind are the most abundant renewable energy sources, but they are also intermittent in nature. This feature raises issues of complementarity and energy storage. Utilities think of generating capacity in terms of baseload, intermediate, and peaking capacities. Because intermittent sources do not fit into any of these categories, they are likely to be relegated to marginal roles.

Inquiries that pose the question of usefulness in another way have produced much more favorable conclusions. Utilities are already accustomed to demand patterns that are subject to seasonal, daily, weather-related, and random fluctuations. Their supply sources, in contrast, can be turned on or off as they wish. The more fruitful approach in considering intermittent sources is to regard them as "negative demand"—the side which is already subject to variations. Models that start with actual hour-by-hour demand loads through the year and then subtract out wind or solar contributions, hour-by-hour, consistently show sizeable potential contributions from intermittent renewables without further backup. Needless to say, the presence of hydroelectric facilities or pumped storage units supports the ability to accommodate intermittent sources.

A simulation of the British electricity system showed that windpower, though not steadily available and not correlated with demand patterns, could still meet 25%–45% of system demand without additional backup, provided that the conventional components could be reconfigured to accommodate wind patterns.[21] A U.S. simulation found that in a system with demand patterns similar to those of most utilities, a grid with a substantial hydroelectric component could accommodate 50% of its outputs from intermittent renewable sources. (i.e., wind and solar) and at costs no higher than those for conventional generation.[22] These are certainly counter-intuitive results, and they would need to be supplemented by modeling studies of other utilities. The findings, though, are consistent with experiences in Denmark and North Germany. Denmark now obtains 21% of its electricity from wind turbines, without evident problems of integrating this output into the national grid. German reports of strains on the grid seem to have more to do with transmission bottlenecks than generation sources.

Solar electricity has the advantage, for most utilities, of coinciding in availability with summer peaks in air-conditioning use. A recent New Jersey study found that photovoltaic generation would not only supply peak power (and thereby displace fuel generation) but 40%–70% of PV capacity would displace the capital costs of conventional generation, fractions which rose quickly in the presence of grid storage capacity.[23]

These results are so much at variance with traditional thinking in utility managements that time, experience, and external pressures will be required to bring solar and wind technologies into widespread use, even when they are cost-competitive. Therein lays the case for incentives and regulatory requirements, such as the renewable portfolio standards now found in 18 states. Expanding markets enable the solar and wind equipment manufacturers to scale up and further reduce costs, while the utilities gain actual operating experience with intermittent sources.

MOVING TOWARD AN ALL-RENEWABLE ENERGY FUTURE

A consideration of energy futures in which renewable sources contribute most or all energy needs begins with an assessment of the quantity of energy that is likely to be needed, and its appropriate forms. As a starting point, one examines present energy sources and end uses in the economy, and then allows for increasing efficiencies. One must show that renewable resources are sufficient in quantity, quality and cost to provide for each of the end uses.

A word of caution is in order. It is much too easy to focus on energy supply, while neglecting the rich possibilities for improving energy efficiency in all end uses. One thus can miss the most economical opportunities for meeting energy demands because efficiency measures usually are less expensive than any new supplies, renewable or conventional. The reader is referred to the many articles in this encyclopedia that deal with the significant potential of increasing the efficiency with which energy is used.

Present U.S. Energy use

U.S. primary energy use is now running at about 100 quads per year and growing at about the same rate as population, unlike the growth pattern prior to 1973. Data for 2003 are shown in Table 1.

As the table indicates, primary energy use in the United States in 2003 was 98.7 quads. About 40% of this primary energy was used to generate electricity, and two-thirds (26.7 quads) of that was lost as waste heat in generation and transmission. End-use energy was thus 72 quads. Transportation and industry each take about one third of

Table 1 Fuel and energy use by sector, United States, 2003 (Quadrillion Btu's)

Sector	Fuel	Electricity	Total end energy use	Electricity losses	Total primary energy	Fuel end uses in sector
Residential	7.2	4.4	11.5	8.7	21.3	Space heat
						Water heat
Commercial	4.3	4.1	8.4	9.2	17.6	Space heat
						Water heat
Transportation	26.8	—[a]	26.8	0.1	26.9	Liquid fuel
Industry						
Manufacturing						
Agriculture, Mining	15.8	3.4	19.2	7.7	26.9	Process heat
Construction	3.0	—[a]	3.0		3.0	
Feed stocks	3.0	—	3.0		3.0	
Total industry	21.8	3.4	25.2	7.7	32.9	
Total all sectors	60.1	11.9	72.0	26.7	98.7	

[a] Less than 0.1 quad.
Source: Adapted from Energy Information Administration (see Ref. 25).

end-use energy, with the rest going to the residential and commercial sectors. In U.S. energy statistics, electricity losses are allocated to each of the sectors in order to see the primary energy demand occasioned by the activities of each sector.

Table 1 also lists the kinds of energy end-uses provided by fuels because these would have to be supplied from renewable sources in the absence of fuels. (Sector end-uses for electricity are not shown, since electricity is readily produced from renewable sources.) In the residential sector, for example, fuels are used mostly to heat space and water or, put differently, to provide low temperature (under 100 degrees centigrade) heat. A much smaller use of fuel is for cooking, which utilizes temperatures in the 100–250 degree centigrade range. Fuel use in the commercial sector is applied for similar purposes. Eighty percent of industrial energy use is in manufacturing, while 20% goes to mining, agriculture, and construction or is used for fuel-based feedstocks. Manufacturing end-use energy is 19 quads of which three to four quads is for refining petroleum.

Energy Efficiency Gains

All of these uses can be reduced with gains in energy efficiency, and some of them can be reduced significantly. Indeed, if the best practices already existing in building design and construction, lights, appliances, heating and cooling systems, vehicle mileage, and industrial processes were uniformly applied in the whole economy, end-use energy demands use would be 50%–70% lower, or in the 22–36 quad range—not the 72 quads shown in Table 1.[24] These potential reductions in energy use make renewable energy futures much more readily conceivable for high-income, high-output nations. Any energy supply system in a world of diminishing oil and gas output would be hard pressed to meet the energy demands of our present inefficient economy.

The Adequacy of Renewable Resources

The questions for renewable energies then become:

1. In the United States, can renewable energy sources provide 22–36 quads, of which six to nine or so would meet efficiency-reduced electricity demands, three to five quads would be used for heat in buildings, seven to ten quads or so would power the transportation sector, and six to ten quads would be composed of fuels for industry?
2. If that can be shown, then can these sources provide for any continuing growth in the economy?
3. What technological developments are needed for this potential to be realized?

The adequacy of renewable energy sources, considered in gross, is easy to establish. Wind and solar resources alone provide these amounts of energy many times over. In addition, the "nonintermittent" renewable sources (biomass, hydroelectric, and geothermal resources) sum to another 17 quads at a minimum.

As to end-uses, six to nine quads of electricity fall well under available sources, with storage capability being required only in areas lacking hydroelectric resources.

Space and water heat can be provided in part from direct solar sources, with the rest coming from biomass or perhaps electricity.

For those who envision a hydrogen-fueled transportation system, intermittent sources of renewable electricity are tailor-made. Hydrogen can be produced by electrolysis whenever solar or wind sources exceed grid demands. Liquid fuels from biomass are likely to be available, as well, especially with emerging technologies to produce liquid fuels from cellulosic materials.

Industrial process heat would come primarily from biomass sources, with some assistance from solar equipment or perhaps hydrogen produced from renewable electricity.

As to growth over time in energy demands over time, one can observe that economic growth in the United States, at least in per capita terms, is directed toward sectors that generally do not have high energy contents (services vs basic commodities). If the United States is ever to reach a measure of sustainability with respect to its physical resource demands, its population must level off, as must that of any other nation seeking sustainability.

There are two major technological developments that remain to be achieved in order to make an all-renewable future attainable. The first is cost reductions in the manufacture and installation of photovoltaic systems and the second is the rapid commercialization and cost reduction in processes which make fuel ethanol from cellulosic materials, rather than foods. These developments are already well underway.

Space does not permit an extended analysis of these matters; this brief discussion is intended to establish, in general terms, that a largely or wholly renewable energy future is possible, given several decades to adjust. This view is not now widely accepted, yet it is certainly the most hopeful one for mankind in an age of diminishing fossil fuel resources.

REFERENCES

1. Kendall, H.W., Nadis, S.J. Eds.; Energy strategies: toward a solar future. In *A Report of the Union of Concerned Scientists*. Ballenger: Cambridge, MA, 1979.
2. Grubb, M.J.; Meyer, N.I. Wind energy: resources, systems, and regional strategies. In *Renewable Energy: Sources for Fuel and Electricity*; Johansson, T.B., Kelly, H., Reddy, A.K.N. et al., Eds.; Island Press: Washington, DC, 1992.
3. Archer, C.L.; Jacobson, M.Z. Spatial and temporal distributions of U.S. winds and windpower at 80 m derived from measurements. J. Geophys. Res. Atmos. **2003**, *108* (D9), 4289.
4. National Hydropower Association, Hydrofacts. http://www.hydro.org/hydrofacts/forecast/asp (viewed May 24, 2005).
5. Wright, P.M. Status and potential of the U. S. geothermal industry. AAPG Bull. **1993**, *77* (8).
6. National Renewable Energy Laboratory, Biomass Energy, Potential, http://www.nrel.gov/documents/biomass_energy.html (viewed May 21, 2005).
7. Perlack, R.D.; Wright, L.L.; Turhollow, A.; Graham, R.L. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*; Oak Ridge National Laboratory: Oak Ridge, TN, April, 2005.
8. Hall, D.O.; Rosillo-Calle, F.; Williams, R.H.; Woods, J. Biomass for energy: supply prospects. In *Renewable Energy: Sources for Fuels and Electricity*; Johansson, T.B., Kelly, H., Reddy, A.K.N., Williams, R.H., Eds.; Island Press: Washington, DC, 1992.
9. Florida Solar Energy Center. http://www.fsec.ucf.edu/solar/testcert/collectr/tpndhw.htm (viewed May 23, 2005).
10. http://www.oja-services.nl/iea-pvps/pv/materials.htm (viewed May 24, 2005).
11. Renewable Energy Policy Project http://www.repp.org./geothermal_brief_power_technologyandgeneration.html (viewed May 24, 2005).
12. http://www.nrel.gov/documents/biomass_energy.html (viewed May 21, 2005).
13. http://www.poemsinc.org/FAQtidal.html (viewed May 25, 2005).
14. Maycock, P.D. *Photovoltaics: Sunlight to Electricity in One Step*; Brick House Publishing Co.: Andover, MA, 1981.
15. http://www.solarbuzz.com Price Survey, May 2005. (viewed May 20, 2005).
16. http://www.nrel.gov/docs/fy04osti/32578.pdf (viewed May 21, 2005).
17. Sawin, J. L. *Mainstreaming renewable energy in the 21st century*. Worldwatch Paper 169, Worldwatch Institute: Washington, DC, 2004. The author gives a figure of 11 cents per kwh in Japan, but this calculation does not include amortization of the equipment or annual maintenance costs.
18. Governors seek wider ethanol use. Wall Street Journal, April 12, 2005.
19. *A New Prosperity; Building a Sustainable Energy Future. The SERI Solar/Conservation Study*. Brick House Publishing Andover, MA, 1981.
20. http://www.renewableenergyaccess.com (viewed May 20, 2005).
21. Grubb, M.J.; Meyer, N.I. Wind energy: resources, systems, and regional strategies. In *Renewable Energy: Sources for Fuels and Electricity*; Johanssen, T.B., Kelly, H., Reddy, A.K.N., Williams, R.H. Eds.; Island Press: Washington, DC, 1992.

22. Kelly, H.; Weinberg, C. Utility strategies for using renewables. In *Renewable Energy: Sources for Fuels and Electricity*; Johanssen, T.B., Kelly, H., Reddy, A.K.N., Williams, R.H. Eds.; Island Press: Washington, DC, 1992.
23. Perez, R. Determination of Photovoltaic Effective Capacity for New Jersey. http://cleanpower.com/research/capacityvaluation/ELCC_New_Jersey.pdf (viewed May 24, 2005).
24. Blackburn, J.O. Possible efficiency gains identified for the U.S. as of the mid-1980's are examined. *The Renewable Energy Alternative*; Duke University Press: Durham, NC, 1987.
25. Energy Information Administration, Monthly Energy Review, March 2005. Additional detail for industry from International Energy Agency, IEA Energy Statistics, 2003, Energy Balance Table for the United States.

Residential Buildings: Heating Loads

Zohrab Melikyan
HVAC Department, Armenia State University of Architecture and Construction, Yerevan, Armenia

Abstract
In designing energy-saving heating systems, it is important to have the exact values of heating loads and seasonal heating demands of buildings. The existing methods for determining these values are not exact enough, because they do not take into account important factors such as the impact of solar radiation on south walls' and roofs' surfaces, the significant difference between day and night outside temperatures, and these temperatures' duration during the heating season. Apart from the stated disadvantage, the existing methods do not take into account the fact that each building has its own duration-of-heating period.

The mentioned imperfections do not allow for providing the exact values of heating demands, which may cause the use of wrong solutions in the designing of heat supply systems.

In this entry, new and improved methods are presented for more precise solution of heating problems.

INTRODUCTION

This entry was prepared based on summarizing the results of a research work, accomplished by the author, within a program of heating efficiency. For providing energy and fuel savings in heating systems, first the exact values of heating loads of buildings are needed. The use of the developed method provides the best possible accuracy in heating-loads value, as it takes into account the impact of more factors. Particularly, this entry highlights that depending on the thermal properties of buildings' envelopes, the duration-of-heating season for each building is different, even in the same climatic conditions. The proposed method is designed for heating experts and it can be used by students.

Method for Heating Load of Building

The heating load is the quantity of heat needed to provide the given inside temperature of a building under the design outside temperature, $t_{\text{out.d}}$ (°C), of given climatic conditions. As a rule, it is more convenient and more exact to calculate the heating load referring to 1 m^3 of a building. This value is called the specific heating demand, q_{hd}, W/m^3, which can be determined by Eq. 1:

$$q_{\text{hd}} = q_{\text{hl}} + q_{\text{v}} + q_{\text{inf}} - q_{\text{d}}, \tag{1}$$

where

- q_{hl} total specific heat lost (W/m^3)
- q_{v} specific quantity of heat required for heating outside fresh air, supplied into the building for ventilation (W/m^3)
- q_{inf} specific quantity of heat lost for heating outside fresh air, penetrating into the building through gaps of windows and doors (W/m^3)
- q_{d} specific quantity of heat gain in the building from lighting, electrical devices, and inhabitants (W/m^3).

For the calculation of the specific values indicated above, we suggest the following equations:

$$q_{\text{hl}} = (t_{\text{in}} - t_{\text{out.dsg}})\left\{2k_{\text{w}}\left(\frac{(1-\mu)}{b} + \frac{1}{a}\right) + \frac{2k_{\text{wd}}\mu}{b} + \frac{k_{\text{r}}}{h}\right\}, \tag{2}$$

$$q_{\text{v}} = 0.181(t_{\text{in}} - t_{\text{out.dsg}}), \tag{3}$$

$$q_{\text{inf}} = \frac{4.012\mu(a+b)(t_{\text{in}} - t_{\text{out.dsg}})}{ab}, \tag{4}$$

By substituting the formulas in Eqs. 2–4 for the sum (Eq. 1), the following equation is obtained for determining the value of specific heating demand for any kind of building:

$$q_{\text{hd}} = (t_{\text{in}} - t_{\text{out.dsg}})\left\{\left[\frac{2k_{\text{wd}}\mu}{b} + \frac{2\left(\frac{1-\mu}{b}+\frac{1}{a}\right)}{\frac{1}{\alpha_{\text{in}}}+\frac{\delta_w}{\lambda_w}+\frac{\delta_{\text{ins}}}{\lambda_{\text{ins}}}+\frac{1}{\alpha_{\text{out}}}}\right] + \frac{1}{h\left(\frac{1}{\alpha_{\text{in}}}+\frac{\delta_r}{\lambda_r}+\frac{\delta_{\text{ins}}}{\lambda_{\text{ins}}}+\frac{1}{\alpha_{\text{out}}}\right)} + 0.181 + \frac{4.012\mu}{b}\right\} - q_{\text{d}}. \tag{5}$$

Keywords: Heating loads; Seasonal loads; Heating designs; Specific heating demand; Design temperature; Heating degree days.

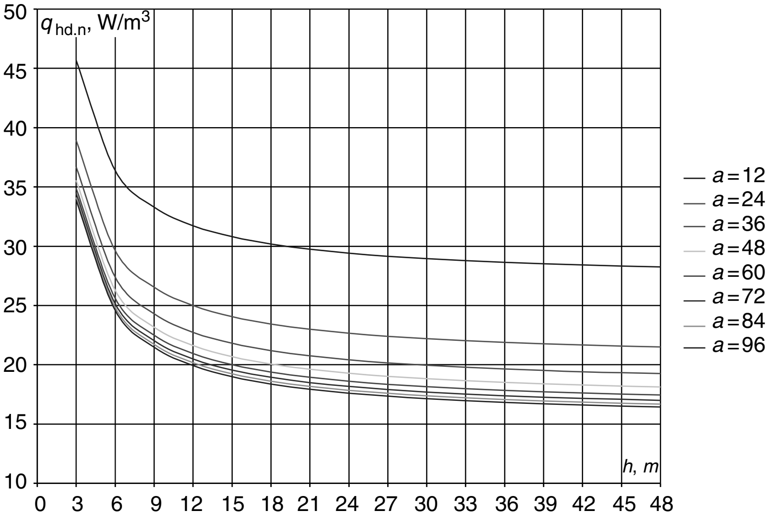


Fig. 1 Values of specific heating demand q_{hd} (W/m^3) depending on sizes *a* and *h*, *m*, of buildings.

In Eq. 5, the meanings of values are the following:

- $t_{out.dsg}$ outside design temperature (°C)
- k_{wd} heat transfer coefficient of windows (W/m^2 °C)
- μ glazing rate of the building
- δ_w and δ_r thickness of the construction material of the walls and ceiling of the building (m)
- λ_w and λ_r heat conductivity of the construction material of the walls and ceiling of the building (W/m °C)
- δ_{ins} thickness of the insulation material covering the walls and ceiling of the building (m)
- λ_{ins} heat conductivity of the insulation material covering the walls and ceiling of the building (W/m °C)
- $\alpha_{in}=8$ W/m^2 °C heat convection coefficient on the inside surfaces of the walls and ceiling of the building
- $\alpha_{out}=20$ W/m^2 °C heat convection coefficient on the outside surfaces of the walls and ceiling of the building.

These values are taken from the building plan and section:

- *a*, *b*, *h* length, width, and height of the building (m)

The glazing rate, μ, of the building is determined by the following fraction:

$$\mu = \frac{\sum F_{wd}}{ah}, \text{ or } \mu = \frac{\sum F_{wd}}{bh},$$

where F_{wd} is the total surface of windows on the vertical surfaces of the building (m^2).

For finding the absolute value of heating demand, Q_{hd}, W, the specific value q_{hd} (W/m^3) is multiplied by the volume, V_b (m^3), of the building:

$$Q_{hd} = q_{hd} V_b \tag{6}$$

The value of internal heat gains, q_d, in the building from lighting, electrical devices, and occupants has considerable impact on the heating and cooling demands of the building. The total value of q_d depends on the electric power of lighting and domestic devices and on the numbers of occupants simultaneously at home. As a result of the analyzed data resulting from a survey of the inhabitants of 1000 apartments in the city of Yerevan under the direction of the author by the methods of mathematical statistics, the following formulas for defining internal heat gains' specific values for winter ($q_{d.w}$) and summer ($q_{d.sum}$) periods have been carried out:

$$q_{d.w} = 0.0016S^3 - 0.0622S^2 + 0.576S + 2.0566 \tag{7}$$

$$q_{d.sum} = 0.0044S^3 - 0.1047S^2 + 0.6551S + 1.9572, \tag{8}$$

where *S* is the number of stories of the building.

The variation of values of specific heating demands, q_{hd}, depending on the length, *a*, and width, *b* (m), of the building is represented in Fig. 1. The diagram shows that the specific heating demands, q_{hd}, for bigger buildings are less, which indicates that the energy efficiency of heating for big buildings is higher.

Method of Seasonal Heating Demand of Buildings

At the present time, the methods for determining the seasonal heating demands of buildings are based on the use of heating degree days or heating degree hours and, therefore, do not take into account the effect of day solar radiation on the surfaces of buildings. Consequently, the values of seasonal heating demands turn out significantly higher. The useful impact of solar radiation on heating demand can be evaluated by the help of radiation temperature, t_R (°C), which is calculated by the following formula:

$$t_R = t_{out} + \frac{Ip}{\alpha_{out}}, \tag{9}$$

where

- t_{out} outside air temperature (°C)
- I intensity of solar radiation on surfaces of building envelope constructions (W/m^2)
- p rate of solar radiation absorption by construction surfaces.

The radiation temperature, t_R (°C), on a given surface is formed in a day period; therefore, it should be evaluated by day outside temperatures, $t_{out.d}$, of the heating period. Based on this approach, the method for determination of seasonal heating demands of buildings is divided into two parts: day, $q_{hd.d.seas}$ (Wh/m^3), and night, $q_{hd.n.seas}$ (Wh/m^3), seasonal heating demands. The total seasonal heating demand will be the sum of daytime and nighttime heating demands:

$$q_{hd.seas} = q_{hd.d.seas} + q_{hd.n.seas}(\text{Wh/m}^3). \tag{10}$$

As the solar radiation intensively lights the south-oriented walls and roof of a building, the heat lost through these constructions for day should be determined by the difference of inside, t_{in}, and radiation, t_R, temperatures. For the constructions of other orientations, heat lost from the building should be determined by the difference of inside, t_{in}, and outside current day temperatures, $t_{out.d.i}$.

Based on this concept, the specific value of total seasonal heat loss through all constructions of the building for the day period is determined by the following equation:

$$q_{hl.d.seas} = \sum_{i=Z_{t_{out.st.d}}}^{Z_{t_{out.dsg}}} Z_{t_{out.d.i}} \left\{ (t_{in} - t_{out.d.i}) \left\{ k_w \left(\frac{2(1-\mu)}{b} + \frac{2}{a} \right) + \frac{k_r}{h} + \frac{2k_{wd}\mu}{b} \right\} - \left[\frac{I_s p_s k_w (1-\mu)}{\alpha_{out} b} - \frac{I_r p_r k_r}{\alpha_{out} h} \right] \right\}, \tag{11}$$

where

- $t_{out.d.i}$ current day temperatures between heating period beginning temperature at day $t_{out.st.d}$ and design temperature $t_{out.dsg}$ of heating period (°C)
- $Z_{t_{out.d.i}}$ duration (h) of each outside day temperature, $t_{out.d.i}$, occurring between day heating season starting temperature, $t_{out.st.d}$, and heating design temperature, $t_{out.dsg}$,
- I_s and I_r average intensity of solar radiation on south wall and roof surfaces (W/m^2).

Besides heat lost, there is direct heat gain, $q_{wd.d.seas}$ (Wh/m^3), through windows by solar radiation. The following formula can be used for determining specific heat gain, $q_{wd.d.seas}$ by day through south windows:

$$q_{wd.d.seas} = \sum_{i=Z_{t_{out.st.d}}}^{Z_{t_{out.dsg}}} Z_{t_{out.d.i}} \frac{\mu I_s n_1 n_2 n_3}{b}, \tag{12}$$

where n_1, n_2, and n_3 are rates of reduction of radiation penetration through windows due to reflection, the window's frames, and dust on glazed surfaces.

By adding the seasonal specific values of $q_{wd.d.seas}$, $q_{v.d.seas}$, $q_{inf.d.seas}$ (Wh/m^3) for heating of ventilation and infiltration air, and substituting the seasonal specific quantity of heat gain, q_d (Wh/m^3), the following equation for determining day seasonal specific heating demand, $q_{hd.d.seas}$ (Wh/m^3) for a building is obtained:

$$q_{hd.d.seas} = \sum_{i=Z_{t_{out.st.d}}}^{Z_{t_{out.dsg}}} Z_{t_{out.d.i}} \left\{ (t_{in} - t_{out.d.i}) \left\{ k_w \left(\frac{2(1-\mu)}{b} + \frac{2}{a} \right) + \frac{k_r}{h} + \frac{2k_{wd}\mu}{b} + 0.181 + \frac{4.012\mu}{b} \right\} - \left[\frac{I_s p_s k_w (1-\mu)}{\alpha_{out} b} + \frac{I_r p_r k_r}{\alpha_{out} h} + \frac{\mu I_s n_1 n_2 n_3}{b} + q_d \right] \right\}. \tag{13}$$

At night period of the heating season, the heat losses through all external constructions of a building take place under the difference of outside night current temperatures, $t_{out.n.i}$ and inside temperature t_{in}. Therefore, the seasonal specific value of heat lost of a building, $q_{hl.n.seas}$ (Wh/m^3) can be determined by the following equation:

$$q_{hl.n.seas} = \sum_{i=Z_{t_{out.st.n}}}^{Z_{t_{out.dsg}}} Z_{t_{out.n.i}} (t_{in} - t_{out.n.i}) \times \left\{ k_w \left(\frac{2(1-\mu)}{b} + \frac{2}{a} \right) + \frac{k_r}{h} + \frac{2k_{wd}\mu}{b} \right\}. \tag{14}$$

By adding the seasonal specific values of $q_{v.n.seas}$ and $q_{inf.n.seas}$ (for heating of ventilation and infiltration air under night temperatures [Wh/m^3]) and substituting the

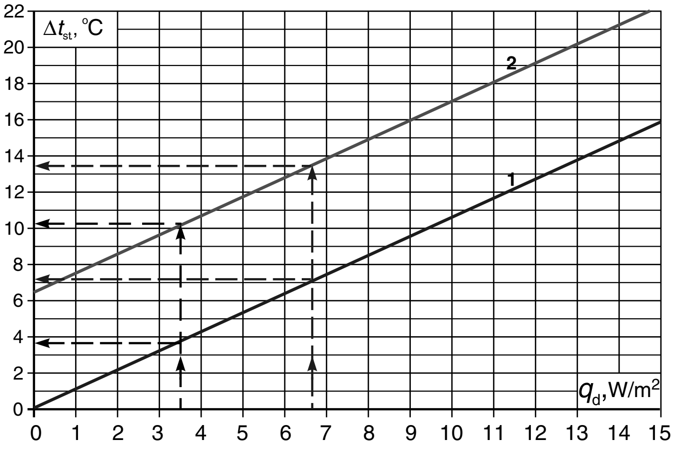


Fig. 2 Difference $\Delta t_{st} = t_{in} - t_{out.st}$ (°C) between inside t_{in} and heating season starting $t_{out.st}$ temperatures in case of different values of internal heat gains, q_d, in a building having sizes $a=48$ m, $b=12$ m, and $h=15$ m (1 for nighttime, 2 for daytime period of heating season).

seasonal specific quantity of heat gain, q_d (Wh/m³), the following equation for determining of night seasonal specific heating demand, $q_{hd.n.seas}$ (Wh/m³) for a building is obtained:

$$q_{hd.n.seas} = \sum_{i=Z_{t_{out.st.n}}}^{Z_{t_{out.dsg}}} Z_{t_{out.n.i}}(t_{in} - t_{out.n.i}) \times \left\{ \left\{ k_w\left(\frac{2(1-\mu)}{b} + \frac{2}{a}\right) + \frac{k_r}{h} + \frac{2k_{wd}\mu}{b} + 0.181 + \frac{4.012\mu}{b} \right\} - q_d \right\}, \quad (15)$$

where

$t_{out.n.i}$ current night temperatures occurring between night heating season's starting temperature, $t_{out.st.n}$, and heating design temperature, $t_{out.dsg}$, (°C)

$Zt_{out.n.i}$ duration (h) of each outside night temperature, $t_{out.n.i}$, occurring between night heating season's starting temperature, $t_{out.st.n}$, and heating design temperature, $t_{out.dsg}$.

To calculate seasonal daytime specific heating demands $q_{hd.d.seas}$, in Eq. 13 is substituted the values of durations $Zt_{out.d.i}$ of daytime current temperatures $t_{out.d.i}$ which occur between temperatures $t_{out.st.d}$ and heating design temperature $t_{out.dsg}$. Seasonal nighttime specific heating demands $q_{hd.n.seas}$, are calculated by substituting in Eq. 15 the values of $Zt_{out.n.i}$ and nighttime current temperatures $t_{out.n.i}$ which occur between temperatures $t_{out.st.n}$ and heating design temperature $t_{out.dsg}$. For this reason, it is necessary to have the durations $Zt_{out.i}$ of each temperature $t_{out.i}$ for each climatic zone, which can be established by special climatologic investigations. For example, for the conditions of Yerevan City, Armenia, the following empirical equations have been obtained:

$$\text{daytime}: \; Z_{t_{out.d.i}} = 0.00003t^5_{out.d.i} + 0.0006t^4_{out.d.i} - 0.0133t^3_{out.d.i} - 0.25t^2_{out.d.i} + 3.92t_{out.d.i} + 61$$

$$\text{nighttime}: \; Z_{t_{out.n.i}} = 0.000006t^5_{out.n.i} + 0.0014t^4_{out.n.i} - 0.031t^3_{out.n.i} - 0.59t^2_{out.n.i} + 9.15t_{out.n.i} + 142$$

For finding out the duration $Zt_{out.d.i}$ (h) of any current daytime temperature $t_{out.d.i}$, within the heating season, the value of that temperature should be substituted in the presented above, first empirical equation. The same kind of procedure should be done for finding out the duration $Zt_{out.n.i}$ (h) of any current nighttime temperature $t_{out.n.i}$, using the second empirical equation.

Heating period's starting daytime $t_{out.st.d}$ and nighttime $t_{out.st.n}$, temperatures are those under which the building is heated enough by internal heat gain, q_d. This means that if the outside temperature is lower than $t_{out.st.d}$ or $t_{out.st.n}$, the building needs to be heated, as the value of heat lost, q_{hl}, becomes higher than q_d.

The values of starting temperatures $t_{out.st.d}$ and $t_{out.st.n}$ can be found analytically, assuming that in Eqs. 5 and 13, the value of current outside temperature, $t_{out.i}$, is equal to $t_{out.st}$. As a result of such a procedure, the following formulas are obtained:

$$t_{out.st.n} = t_{in} - \frac{q_d}{2\left(\frac{k_{wd}\mu}{b} + k_w\left(\frac{1-\mu}{b} + \frac{1}{a}\right)\right) + \frac{k_r}{h} + 0.181 + \frac{4.012\mu}{b}} \quad (16)$$

$$t_{out.st.d} = t_{in} - \frac{\frac{I_s p_s k_w(1-\mu)}{\alpha_{out} b} + \frac{I_r p_r k_r}{\alpha_{out} h} + \frac{\mu I_s n_1 n_2 n_3}{b} + q_d}{2k_w\left(\frac{(1-\mu)}{b} + \frac{1}{a}\right) + \frac{k_r}{h} + \frac{2k_{wd}\mu}{b} + 0.181 + \frac{4.012\mu}{b}} \quad (17)$$

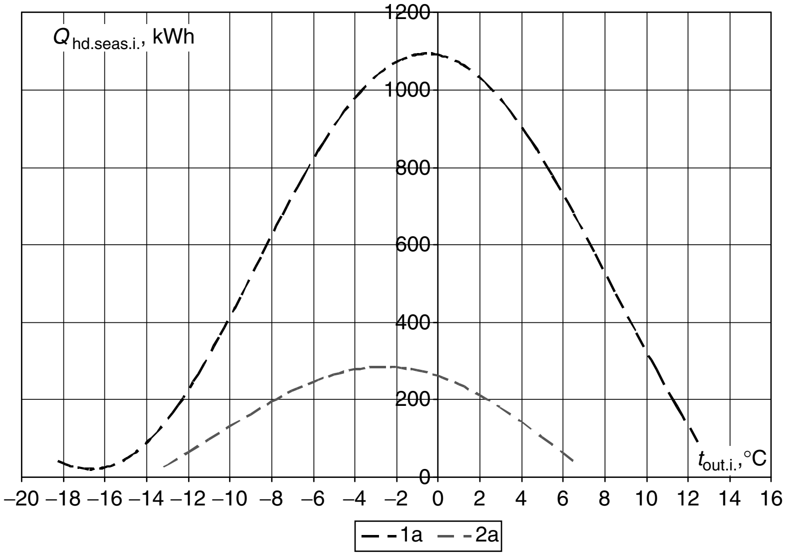


Fig. 3 Daytime and nighttime seasonal specific heating demands for an individual house in the climatic conditions of Yerevan, Armenia; (1a) nighttime seasonal specific heating demands, $q_{hd.n..i}$ (kWh/m^3); (2a) daytime seasonal specific heating demands, $q_{hd.d..i}$ (kWh/m^3).

The Eq.16 permits to find the temperature $t_{out.st.d}$, under which starts the nighttime heating season for any kind of building. By the Eq. 17 the temperature $t_{out.st.d}$, under which starts the daytime heating season, is determined. The impact of internal heat gains q_d,W, on the heating seasons' starting $t_{out.st.d}$, and $t_{out.st.d}$, temperatures is significant, which can be seen from the diagram in Fig. 2. The diagram shows that the growth of q_d, increases the difference $\Delta t_{st} = t_{in} - t_{out.st}$, between inside t_{in} and heating season's starting temperature $t_{out.st}$. This means that in the buildings with higher internal heat gains q_d, the heating starts at lower outside temperatures, and lasts shorter.

Using the collection of methods provided above, simulation software has been developed that calculates the exact values of seasonal heating demands for residential buildings.

The results of the calculations of day and night seasonal heating demands for individual house in Yerevan climatic conditions are represented by Fig. 3.

The diagram proves that for the examined individual house, the night seasonal specific heating demand, $q_{hd.n..i}$, becomes zero ($q_{hd.n..i} = 0$) under outside temperature $t_{out.n.i} < +15$°C. The day seasonal specific heating demand, $q_{hd.d..i}$, becomes zero ($q_{hd.d..i} = 0$) under outside temperature $t_{out.d.i} < +8$°C. This means that the heating of the house at night should start when the outside temperature $t_{out.n.i} < +15$°C, and at day it should start when the outside temperature $t_{out.d.i} < +8$°C.

The sum of all day and night current heating demands under all possible outside temperatures represents the total seasonal specific heating demand for the house. Multiplying total seasonal specific heating demands by the volume of the house will obtain the absolute value of seasonal heating demand.

The experimental calculations of seasonal heating demands for the same building by existing and suggested methods prove that the correct value of heating demand, defined by the help of new method, in 25%–30% is less, than in case of calculation by existing method. This difference is explained by consideration of solar energy impact and by more correct approach to the determination of heating degree hours for each building.

CONCLUSIONS

1. The suggested methods in this entry can be applied for accurate calculation of seasonal heating demands of any kind of building that possess any combination of thermal and physical properties of constructions in any climatic conditions. For this purpose, it is necessary to conduct climatologic investigations and develop empirical equations for determining the total duration of current daytime and nighttime temperatures in given climatic conditions.
2. Each building has its own heating season's beginning temperature, regardless of climatic conditions. The same building in various climatic conditions has the same heating season starting temperature, but the heating season period of a building depends on the total duration of temperatures having values between the heating season starting temperature, $t_{out.st.n}$, and the heating design temperature, $t_{out.dsg}$, in a given area.

3. The accurate values of seasonal specific heating demands allow determining and planning accurate values of fuel consumption for heating purposes.

BIBLIOGRAPHY

1. Turner, W.C., Ed. *Energy Management Handbook* 6th Edition; Fairmont Press: Atlanta, GA, 2006.
2. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. *ASHRAE Handbook of Fundamentals*; ASHRAE: Atlanta, GA, 2005.
3. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. *HVAC Systems and Equipment*; ASHRAE: Atlanta, GA, 2004.
4. Goldberg, N. Air conditioning heating and ventilating. *Marks' Standard Handbook for Mechanical Engineers* 10th Ed.; McGraw Hill: New York, 1996; [Chapter 12.4].
5. Sawer, H.J.; Howell, R.H. *Principles of Heating, Ventilating and Air Conditioning*; ASHRAE: Atlanta, GA, 1994.

Run-Around Heat Recovery Systems

Kamiel S. Gabriel
University of Ontario Institute of Technology, Oshawa, Ontario, Canada

Abstract
Run-around heat-recovery systems are often used to recover heat from the exhaust air in building ventilation systems, particularly in cold climates. In a typical heat-recovery system, an ethylene glycol and water solution is used as a "coupling fluid" to prevent the system from freezing. The design of a run-around heat-recovery system involves consideration of the heat-transfer rates between the fluids. A challenging problem is how to significantly increase those rates between the fluids, while using less energy for heating and cooling loads in buildings. Due to the heat-transfer characteristics of this coupling fluid and the operating and capital cost factors, the typical overall effectiveness of such systems is only about 50%. Studies show that two-phase, gas–liquid coupling fluids have much higher convective heat-transfer rates than single-phase flows at the same mass flow rates.

INTRODUCTION

In industrialized countries, a large portion of the energy consumed is discharged into the ambient. The mining and manufacturing industries account for a large percentage of waste heat. In particular, the iron and steel industry produces a large amount of waste energy with a broad range of temperatures.

In order to reduce the consumption of oils and fossil fuels, emphasis has been placed on the utilization of alternative energy sources, as well as energy conservation. The latter could be achieved through the enhancement of heat-transport systems (thus, reducing the production of waste heat), or by utilizing waste heat that has been already produced to pre-heat incoming air into buildings. Waste energy recovery can be achieved in many ways. Some examples are:

1. Using a pre-heater to transfer heat to another working fluid
2. Conversion into other forms of energy (e.g., electrical generators using steam)
3. Heat storage for control of supply-demand cycles between the source of waste heat and the consumer
4. Using run-around heat-recovery systems employing coupling fluids that absorb heat from the source (waste energy) and transfer it back into usable space

This entry will focus on the last method: run-around heat-recovery systems.

Keywords: Heat-recovery systems; Energy conservation; Heat-exchangers; Gas–liquid flow.

RUN-AROUND HEAT-RECOVERY SYSTEMS

The technique of waste heat-recovery involves heat transfer from the hot fluid (from which the energy is to be extracted) to the cold fluid (which is the energy absorbent). This system is comprised of two air-to-liquid heat exchangers, or coils, that are thermally connected by a pumped liquid. In the Heating, Ventilation, and Air Conditioning (HVAC) industry, an ethylene glycol and water solution is used as the "coupling fluid" to prevent the system from suffering freeze damage in cold winter conditions. This system is favorable for applications that involve remote air streams or require complicated configurations of air ducting, such as those found in the heating and ventilating systems of large buildings.

Due to the heat-transfer characteristics of this coupling fluid and the operating and capital cost factors, the typical overall effectiveness of such systems is only about 50%.[1] The performance of the run-around system depends on many parameters, such as coil geometry, air flowrates, air temperatures, fan power requirements, glycol flowrates, pump power requirements, heat losses in the system, and operating schedules. Computer simulations were conducted by several researchers to investigate the system's performance.[2–4] The results demonstrate that a major change in the coupling fluid is needed to facilitate significant improvement in the overall performance of the system.

Studies related to two-phase, gas–liquid flows in tubes have shown that the heat-transfer coefficients associated with mixtures of liquids and vapors (or gases) have much higher values than those with single-phase flow at the same mass flowrates.[5–7] Rezkallah and Sims[7] developed a general correlation for heat transfer in two-phase, gas–liquid flow in a vertical tube. However, there had been little or no heat transfer data available in a tube configuration similar to that of a typical heat exchanger coil until the work published by Zeng, Rezkallah, and Besant.[8] In addition to obtaining a large set of heat-transfer and pressure-drop data about typical heat-exchanger coils on the supply and exhaust sides of a run-around heat-recovery loop, the researchers[8] also examined the effect of temperature-dependent properties on the performance of the heat-recovery system.[9]

The results have shown that the temperature-dependent properties may further alter the performance of run-around systems if the coupling fluid experiences large temperature cycles in the run-around loop. Since run-around heat-recovery systems using liquid aqueous-glycol coupling fluids are usually designed and operated with low overall heat-transfer units, high Reynolds numbers must be maintained and the maximum overall system effectiveness is limited to 50%.[1] Therefore, pumping rates in a run-around loop need to be carefully selected, taking into account the glycol concentration and the supply air inlet temperature, if the very significant effects of temperature-dependent properties are to be avoided.[9] If, on the other hand, coupling fluids with freeze protection and much better heat-transfer characteristics can be found, the number of transfer units can be raised, resulting in a much higher maximum overall effectiveness. One fluid that has this potential is a liquid–gas mixture described by Zeng, Rezkallah, and Besant.[8]

Fig. 1 shows a schematic of a typical heat-recovery system. It consists of two major subsystems; an air supply system and a circulating fluid (usually known as the "coupling fluid"). In cold climates, the coupling fluid is typically an aqueous glycol solution to prevent the system from damage by freezing. In both the supply and exhaust ducts, air is supplied by fans. The air flows through a normal duct system. For the best performance of the system, the air flowrates (supply and exhaust) need to be balanced. This can be monitored using air pressure transducers.

Res–Solar

HEAT-TRANSFER COEFFICIENTS

The overall thermal conductance, UA, for each coil is calculated from:

$$\frac{1}{UA} = \frac{1}{(\eta hA)_{\mathrm{h}}} + R_{\mathrm{sh}} + R_{\mathrm{sc}} + R_{\mathrm{w}} + R_{\mathrm{c}} + \frac{1}{(\eta hA)_{\mathrm{c}}} \tag{1}$$

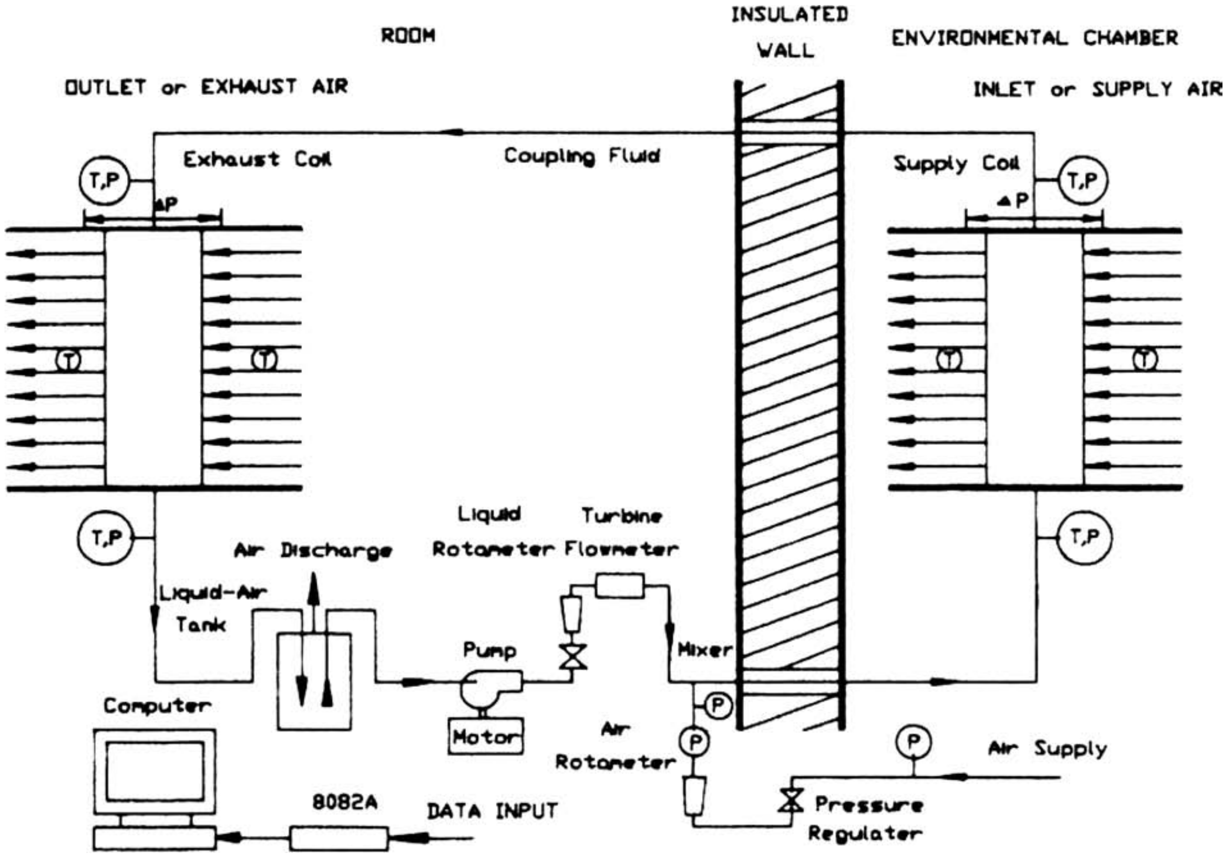


Fig. 1 A schematic of a typical run-around heat-recovery system.

where U is the overall heat-transfer coefficient (W/m^2 K); A is the tube cross-section area; c and h refer to the cold and hot fluids, respectively; R_{sh} and R_{sc} are the fouling resistances on the hot and cold sides, respectively; and R_w and R_c are the total tube wall thermal resistance and contact resistance of fins and tube, respectively. All resistances are in K/W.

A parameter known as the number of transfer units, N, is commonly used in the equations for calculating the effectiveness of a coil from:

$$N = \frac{UA}{C_{min}} \tag{2}$$

The heat capacity ratio, Cr, is estimated from:

$$Cr = \frac{C_{min}}{C_{max}} \tag{3}$$

where C_{min} and C_{max} are the minimum and maximum heat capacities of the cold and hot fluids (W/K).

Fig. 2 is a plot of the thermal effectiveness, ε_s, against the number of transfer units for heat capacity ratios ranging from 0.58 to 0.97. The experimental data of Zeng, Rezkallah, and Besant[8] are also shown on the plot.

The heat-transfer rate, q, is calculated from:

$$q = \varepsilon q_{max} = UA(\text{LMTD})F = C_h(dT_h)$$
$$= C_c(dT_c) \tag{4}$$

where F is used as a correlation factor with the Logarithmic Mean Temperature Difference (LMTD) method,[10] and is equal to 1.0 in most cases. From Eq. 4, UA can be calculated and the heat transfer coefficients for single-phase liquid flow (h_L), and two-phase flow (h_{TP}) can then be obtained.

Since the minimum heat capacity in the system is that of the supply air, no dramatic changes could be made to the value of N. However, the number of transfer units in each coil and, hence, the overall effectiveness of the run-around system can be significantly increased if the overall heat transfer coefficient U is increased. Fig. 3 is a plot of the two-phase heat-transfer coefficients against the gas quality, x (defined as the ratio of the air mass flow rate to the total mass flow rate of the water–air mixture), for the supply and exhaust coils at liquid flow Reynolds numbers of 7700 and 12,700. It can be easily observed that as the air quality increases, the heat-transfer coefficients increase, yielding a better overall thermal effectiveness of the run-around system. Increasing the liquid flow rate (or Reynolds number) also produces higher heat-transfer coefficients. The heat-transfer coefficient for the supply coil increased by approximately 100% (at $x = 0.004$) when the Reynolds number was increased by 65%. It can be also concluded from the results in Fig. 3 that the exhaust coil heat-transfer coefficient is larger than that of the supply coil.

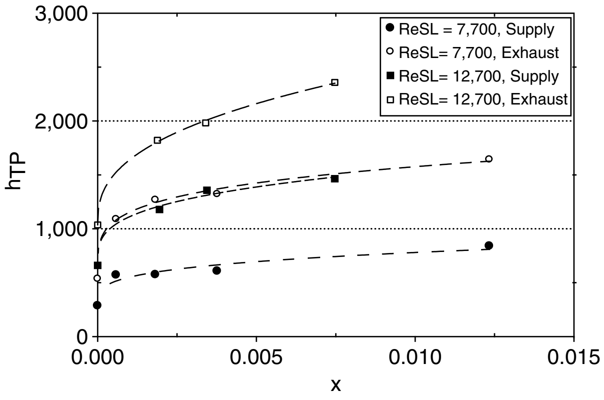


Fig. 3 Two-phase heat transfer coefficients as a function of gas quality using water–air coupling fluid, air supply volumetric flow rate; $Q_{as} = 21$ m^3/min. air exhaust volumetric flow rate; $Q_{ae} = 22$ m^3/min.

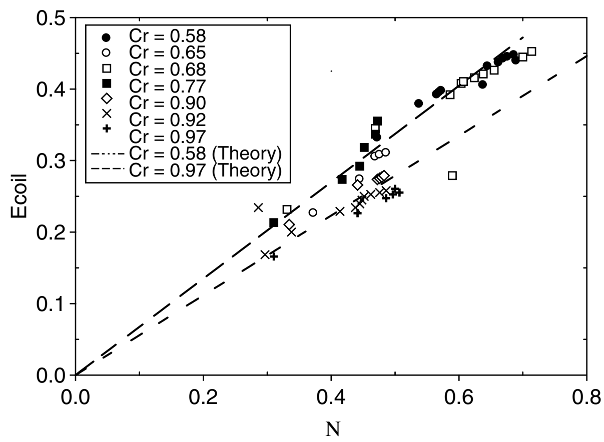


Fig. 2 Coil effectiveness as a function of number of transfer units with Cr as a parameter.

INFLUENCE OF INLET SUPPLY AIR TEMPERATURE

In general, higher heat-transfer coefficients are associated with lower inlet supply air temperatures. As shown in Fig. 4, the relationship tends to be linear.

INFLUENCE OF AIR FLOW RATE THROUGH THE DUCT

Fig. 5 shows two sets of data for the heat-transfer rate for 35% glycol/water–air coupling fluid at two different duct air flow rates. The results are those of Zeng, Rezkallhah, and Besant.[8] It can be observed that higher heat-transfer rates are always associated with lower duct air flow rates;

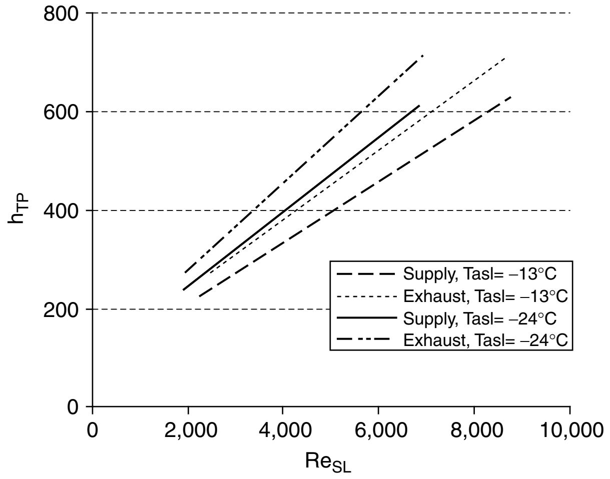


Fig. 4 Two-phase heat transfer coefficients as a function of the liquid Reynolds number using 35% glycol/water–air coupling fluid with $x = 0.0007$, Air supply volumetric flow rate; $Q_{as} = 21\ m^3/min$, Air exhaust volumetric flow rate; $Q_{ae} = 22\ m^3/min$.

e.g., at $Re_{SL} = 2500$, h_{TP} decreased by 16% when the duct air flow rates were increased. This is partly attributed to the significant change in viscosity of the glycol/water solution as a result of the change in operating temperature.

INFLUENCE OF CHANGES IN THE OPERATING TEMPERATURE OF ETHYLNE GLYCOL

Experimental and numerical studies were performed by Zeng, Besant, and Rezkallah[9] to examine the performance of run-around systems when the coupling fluid properties vary with temperature. For a 50% aqueous glycol solution, it was shown that the temperature variations in the glycol solution produced small changes in the supply and exhaust air temperatures. Both the supply and exhaust air temperatures have shown a non-linear variation with position along the coil tube. In the case where properties were independent of the coupling fluid temperature, the relationship was linear. These results are shown in Fig. 6.

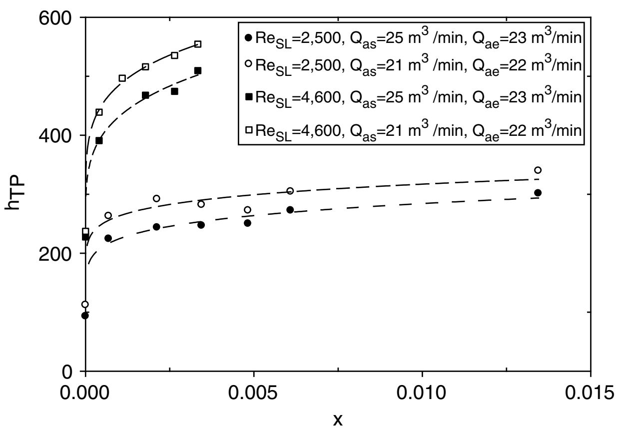


Fig. 5 Two-phase heat transfer coefficients as a function of gas quality using 35% glycol/water–air coupling fluid, Air supply volumetric flow rate; $Q_{as} = 21\ m^3/min$, Air exhaust volumetric flow rate; $Q_{ae} = 22\ m^3/min$.

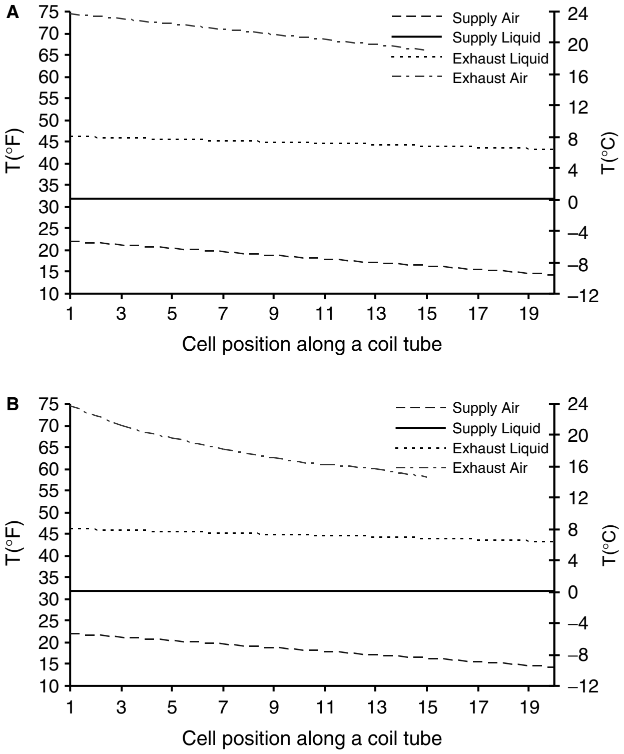


Fig. 6 Temperature distribution in a run-around system (A) Properties independent of coupling fluid temperature (B) Properties dependent on the local coupling fluid temperature.

OPERATING FEATURES OF THE TWO-PHASE FLOW RUN-AROUND SYSTEM

(Refer to Fig. 1 for the component identification)

The coupling fluid is pumped around the run-around system at a rate controlled by the flow control by-pass valve using a pump(s) sized to provide a coupling liquid flow rate such that the thermal capacity ratio for each coil is equal to one when no flow is by-passed.

The coupling liquid flow is divided equally to each tube in each coil using flow divider valves (not shown in the figure).

Pressurized gas is injected into the coupling liquid flow in each tube of each coil in such a manner that the gas flow is distributed equally in each tube. This rate of gas injection will result in a mass ratio of at least 0.0001 between the gas and the coupling liquid.

The rate of coupling liquid flow and gas flow in each coil is adjusted to maximize the total heat exchange between coils or, in the event that the required load is met, to reduce the total heat exchange between the coils.
The mixed gas and coupling liquid is separated after they flow through each coil.

CONCLUSIONS

Run-around heat-recovery systems are very effective means of reducing the heating costs of buildings during winter months, particularly for northern climates, such as Canada. A recovery system can also reduce the size and capital cost of supporting HVAC equipment, such as boilers and radiators. A practical system for recovering this lost energy is a run-around, or liquid-coupled, heat recovery system.

Heat transfer rates in run-around heat-recovery systems of the type commonly used in HVAC systems are greatly enhanced when two-phase, gas–liquid flow through the coils is utilized. The overall effectiveness of the system increases when the two-phase mixture replaces the commonly used single-phase aqueous glycol solution. Moreover, by using the liquid–gas coupling fluid, the capacity of the coupling fluid could be matched to that of the exhaust and supply air. This leads to a substantial enhancement of the heat transfer process in such systems, resulting in higher efficiency and revenue saving.

REFERENCES

1. Forsyth, B.I.; Besant, R.W. The design of a run-around heat-recovery system. ASHRAE Trans. **1988**, *94* (II), 511–531.
2. Custer, J.B. The coil energy recovery loop. ASHRAE Trans. **1976**, *88*, 998–1007.
3. Wang, J.C.Y. *Practical Thermal Design of Run-Around Systems*, Proceedings of the Solar Energy Society of Canada Energex 8th Conference, ISES: Regina, Canada, 1982; 641–645.
4. Forsyth, B.I.; Besant, R.B. The performance of a run-around heat-recovery system using aqueous glycol as a coupling liquid. ASHRAE Trans. **1988**, *94* (II), 532–545.
5. Collier, J.B. *Convective Boiling and Condensation*, 1st Ed.; McGraw Hill Company: U.K., 1972; 382–406.
6. Michiyoshi, I. *Two-Phase Two-Component Heat Transfer*, Proceedings of the 6th International Heat Transfer Conference, Washington, DC, 1978, 1, 219–233.
7. Rezkallah, K.S.; Sims, G.E. An examination of correlations of mean heat-transfer coefficients in two-phase two-component flow in vertical tubes. AICHE Symp. Ser. **1987**, *83*, 109–114.
8. Zeng, Y.Y.; Rezkallah, K.S.; Besant, R.W. Heat transfer measurements in two-phase two-component flow inside heat exchanger coils of a run-around heat recovery system. *ASME Publication HTD*; Two-phase flow in energy exchange systems; 1992, 1–10.
9. Zeng, Y.Y.; Besant, R.W.; Rezkallah, K.S. The effect of temperature-dependent properties on the performance of run-around heat recovery systems using aqueous-glycol coupling fluid. ASHREA Trans. **1993**, 551–562.
10. Incropera, F.P.; DeWitt, D.P. *Fundamentals of Heat and Mass Transfer*, 5th Ed.; Wiley: U.S.A., 2002.

Savings and Optimization: Case Studies

Robert W. Peters
Department of Civil, Construction & Environmental Engineering, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.

Jeffery P. Perl
Chicago Chem Consultants Corp., Chicago, Illinois, U.S.A.

Abstract
Since mass and energy balances are coupled, process energy can be optimized only by conducting detailed energy assessments, which investigate both process and energy analyses. A number of plant-wide assessments have been conducted on a variety of industries, including aluminum, chemicals, forest products, glass, metal casting, mining, petroleum, and steel, among others. While a number of industrial organizations have had plant-wide assessments performed at their facilities in each of the above industries [reported on the DOE website], one industrial organization was selected from each type of industry reported above, and the results and recommendations of the energy assessments are described in this chapter as case histories. Examples of recommendations from the industrial energy assessments include: improve waste heat recovery, consolidation of cooling/chiller operations, improved process recycling operations, lower temperature operation, use of variable speed drives on process equipment, implementing more efficient plant lighting systems, reducing scrap/wastage rates, and upgrading equipment with more energy efficient systems. Potential energy savings in terms of payback periods, capital savings, and reductions of Btu's and kWh/yr are reported, for the various assessment recommendations presented.

The U.S. Department of Energy (DOE) has an industrial technology program (ITP) support available. Elements of this program include free technical assistance as well as project co-funding opportunities. The program identifies four broad areas in the Save Energy Now arena and can provide free energy audits to facilities around the US. These audits assist companies in charting facilities energy use, identify areas of concern by matching energy bills with similar facilities and recommending changes along with estimated pay back periods to allow project prioritization

Although this entry is meant to complement our preceding article "Energy Savings and Optimization in the Chemical Process Industry" in this encyclopedia, the examples also apply to a wide variety of applications wherever energy is used in any manufacturing process.

U.S. DOE CASE STUDIES

DOE has published a number of best practices of industrial technology energy optimization case studies at their website: www1.eere.energy.gov/industry/bestpractices/printable_versions/case_studies.html. The write-ups are based on the case studies provided in the DOE Energy Efficiency and Renewable Energy (EERE) website: www1.eere.energy.gov/news/progress_alerts/progress_alert.asp?aid=183. The following case studies involving plant-wide assessments are available at that website, divided according to industrial category:

Aluminum

- Alcoa: Lafayette Operations Energy Efficiency Assessment (PDF 235 KB);
- Alcoa: Plant-Wide Energy Assessment Finds Potential Savings at Aluminum Extrusion Facility (PDF 464 KB);
- Commonwealth Aluminum: Manufacturer Conducts Plant-Wide Energy Assessments at Two Aluminum Sheet Production Operations (PDF 810 KB);
- Pechiney Rolled Products: Plant-Wide Energy Assessment Identifies Opportunities to Optimize Aluminum Casting and Rolling Operations (PDF 417 KB).

Chemicals

- 3M: Hutchinson Plant Focuses on Heat Recovery and Cogeneration during Plant-Wide Energy-Efficiency Assessment (PDF 710 KB);

Keywords: Case studies; Energy efficiency; Payback period; Energy efficiency assessment; Plantwide assessment; Energy optimization; Energy savings; Process system analysis; Waste heat recovery; Variable-speed drive.

- Akro Nobel Morris Plant Implements a Site-Wide Energy Efficiency Plan (PDF 341 KB);
  - — Plant–Wide Assessment (PWA) Summary: $1.2 Million in Savings Identified on Akro Novel Assessment (PDF 174 KB);
- Bayer Polymers: Plant Identifies Numerous Projects Following Plant-Wide Energy Efficiency Assessments (PDF 863 KB);
- Dow Chemical Company: By-Product Synergy Process Provides Opportunities to Improve Resource Utilization, Conserve Energy, and Save Money (PDF 218 KB);
- Formosa Plastics Corporation: Plant-Wide Assessment of Texas Plant Identifies Opportunities for Improving Process Efficiency and Reducing Energy Costs (PDF 1 MB);
- Neville Chemical Company: Management Pursues Five Projects Following Plant-Wide Energy-Efficiency Assessment (PDF 1.0 MB);
- Rohm and Haas: Chemical Plant Uses Pinch Analysis to Quantify Energy and Cost Savings Opportunities at Deer Park, Texas (PDF 607 KB);
- Rohm and Haas: Company Uses Knoxville Plant Assessment Results to Develop Best Practices Guidelines and Benchmark for Its Other Sites [Revised July 2003] (PDF 765 KB);
- Solutia: Massachusetts Chemical Manufacturer Uses SECURE Methodology to Identify Potential Reductions in Utility and Process Energy Consumption (PDF 979 KB); and
- W.R. Grace: Plant Uses Six Sigma Methodology and Traditional Heat Balance Analysis to Identify Energy Conservation Opportunities at Curtis Bay Works (PDF 906 KB).

Forest Products

- Appleton Papers Plant-Wide Energy Assessment Saves Energy and Reduces Waste (PDF 220 KB);
  - — PWA Summary: $3.5 Million in Savings Identified in Appleton Assessment (PDF 176 KB);
- Augusta Newsprint: Paper Mill Pursues Five Projects Following Plant-Wide Energy Efficiency Assessment [Revised July 2003] (PDF 699 KB);
  - — PWA Summary: $1.6 Million in Savings Identified in Augusta Newsprint Assessment (PDF 176 KB);
- Blue Heron Paper Company: Oregon Mill Uses Model-Based Energy Assessment to Identify Energy and Cost Savings Opportunities (PDF 394 KB);
- Boise Cascade Mill Energy Assessment (PDF 297 KB);
  - — PWA Summary: $707,000 in Savings Identified in Boise Cascade Assessment (PDF 176 KB);
- Caraustar Industries Energy Assessment (PDF 188 KB);
  - — PWA Summary: $1.2 Million in Savings Identified in Caraustar Assessment (PDF 176 KB);
- Georgia-Pacific Palatka Plant Uses Thermal Pinch Analysis and Evaluates Water Reduction in Plant-Wide Energy Assessment (PDF 352 KB);
  - — PWA Summary: $2.9 Million in Savings Identified in Georgia-Pacific Assessment (PDF 176 KB);
- Georgia-Pacific: Crossett Mill Identifies Heat Recovery Projects and Operational Improvements That May Save $6 Million Annually (PDF 797 KB);
- Inland Paperboard and Packaging, Rome Linerboard Mill Energy Assessment (PDF 280 KB);
  - — PWA Summary: $9.5 Million in Savings Identified through Inland Assessment (PDF 176 KB);
- Weyerhaeuser Company: Longview Mill Conducts Energy and Water Assessment that Finds Potential for $3.1 Million in Annual Savings (PDF 400 KB).

Glass

- Corning, Inc.: Proposed Changes at Glass Plant Indicate $26 Million in Potential Savings (PDF 177 KB); and
- Anchor Glass Container Corporation: Plant-Wide Energy Assessment Saves Electricity and Expenditures (PDF 200 KB);
- PWA Summary: $1.6 Million in Savings Identified in Anchor Assessment (PDF 181 KB).

Metal Casting

- AMCAST Industrial Corporation: Energy Assessment (PDF 295 KB);
  - — PWA Summary: $3.6 Million in Savings Identified in AMCAST Assessment (PDF 176 KB);
- Ford Cleveland: Inside-Out Analysis Identifies Energy and Cost Savings Opportunities at Metal Casting Plant (PDF 266 KB).

Mining

- Alcoa World Alumina: Plant-Wide Assessment at Arkansas Operations Reveals More than $900,000 in Potential Annual Savings (PDF 464 KB);
- Coeur Rochester, Inc.: Plant-Wide Assessment of Nevada Silver Mine Finds Opportunities to Improve

Process Control and Reduce Energy Consumption (PDF 300 KB); and
- Kennecott Utah Copper Corporation: Facility Utilizes Energy Assessments to Identify $930,000 in Potential Annual Savings (PDF 307 KB).

Petroleum

- Chevron: Refinery Identifies $4.4 Million in Annual Savings by Using Process Simulation Models to Perform Energy-Efficiency Assessment (PDF 293 KB);
- Martinez Refinery Completes Plant-Wide Energy Assessment (PDF 241 KB); [Equilon]
  — PWA Summary: $52 Million in Savings Identified in Equilon Assessment (PDF 172 KB);
- Paramount Petroleum: Plant-Wide Energy-Efficiency Assessment Identifies Three Projects (PDF 816 KB);
  — PWA Summary: $4.1 Million in Savings Identified in Paramount Petroleum Assessment (PDF 173 KB); and
- Valero: Houston Refinery Uses Plant-Wide Assessment to Develop an Energy Optimization and Management System (EOMS) (PDF 308 KB).

Steel

- Jernberg Industries, Incorporated: Forging Facility Uses Plant-Wide Energy Assessment to Aid Conversion to Lean Manufacturing (PDF 987 KB);
- Full PWA Report: An Assessment of Energy, Waste, and Productivity Improvements for North Star Steel Iowa (PDF 2.32 MB);
- North Star Steel Company: Iowa Mini-Mill Conducts Plant-Wide Energy Assessment Using a Total Assessment Audit (PDF 608 KB); and
- Weirton Steel: Mill Identifies $1.4 Million in Annual Savings Following Plant-Wide Energy-Efficiency Assessment (PDF 555 KB).

Supporting Industries

- Metaldyne: Plant-Wide Assessment of Royal Oaks Finds Opportunities to Improve Manufacturing Efficiency, Reduce Energy Use, and Achieve Significant Cost Savings (PDF 676 KB).

Other

- Aluminum/Alcoa: IAC Energy Assessment of Spanish Fork Plant (PDF 174 KB); and
- Metlab: Plant-Wide Assessment (PDF 640 KB);
- Utica Corporation: Plant-Wide Energy Assessment Report Final Summary (PDF 225 KB).

Examples of a plant energy optimization assessment from these various major industries are summarized below.

Aluminum

Commonwealth Aluminum (now Aleris Rolled Products): Commonwealth Aluminum conducted plant-wide energy assessments at its aluminum sheet rolling mills in Lewisport, Kentucky, and Uhrichsville, Ohio. The Lewisport facility utilizes direct-chill ingot castings, and the Uhrichsville facility employs continuous casting technology. The direct-chill process makes use of reversing hot mills to homogenize, scalp, and break down large cast ingots. Additional processing includes hot reduction using multistand mills, coiling, annealing, cold reduction, and finishing. The continuous-casting operation has molten metal fed from an adjacent scrap casting operation and from melted in-house scrap to the continuous-casting unit, producing an aluminum strip. The as-cast strip is fed to a multistand hot mill and then goes to coiling, annealing, cold reduction, and finishing operations; this process eliminates the scalping, homogenization, and reversing hot mill operations that are part of the direct-chill process.

The assessments focused on analyzing the processing procedures involved in converting scrap feed material to finished coiled sheet products. Specifically, the assessments involved looking at the role of process technologies; the impact of operating procedures; and the effects of alloy composition, temper, sheet width, sheet thickness, and coil weight and size to determining energy requirements. The assessment goals were: (i) identify levels of energy consumption inherent in the direct-chill and continuous casting processes; (ii) identify the best practices at each plant and recommend potential methods to transfer those practices; and (iii) define opportunities for improving future energy consumption by investing in new equipment, changing operating practices, and optimizing production strategies. The assessment team reviewed: (1) process heating, with an emphasis on melting, homogenization, annealing, and painting processes; (2) motors and pumps (including those used for hot and cold rolling operations), saws and shears, material-handling equipment, and environmental control systems; (3) compressed air systems (including locating and repairing air leaks); (4) steam systems, with an emphasis on steam usage for process and plant heating and cooling; and (5) application of new processes and equipment, including energy contracting and scheduling, waste heat recovery, lighting, and heating, ventilating, and air-conditioning equipment. Ten potential projects were identified by this assessment

Res–Solar

Table 1 Projects identified from Commonwealth Aluminum's Plantwide Assessment

Project	Fuel savings (MMBtu/yr)	Electricity savings (kWh/yr)	Cost savings ($/yr)	Capital cost ($)	Payback period (years)
	Annual projected savings				
1. Convert the Lewisport mill to the continuous-casting process	546,000	30,900,000	8,000,000	25,000,000	3.1
2. Upgrade melter/holder furnaces at Newport mill	153,000	N/A	918,000	5,000,000	5.4
3. Modify the production strategy to accommodate market demands	83,000	N/A	500,000	None	Immediate
4. Implement the best practices for melting operations	46,000	N/A	273,000	None	Immediate
5. Improve melt stirring processes at Newport mill	33,000	N/A	200,000	400,000	2.0
6. Upgrade soaking pits at Lewisport mill	128,000	N/A	770,000	500,000	0.6
7. Improve annealing operations	22,000	N/A	130,000	None	Immediate
8. Optimize compressed air system	N/A	6,800,000	238,000	248,000	1.0
9. Use infrared imaging technology for process diagnostics	50,000	N/A	300,000	None	Immediate
10. Improve waste heat recovery	Unknown	Unknown	Unknown	Unknown	Unknown
Total	1,061,000	37,700,000	11,329,000	31,148,000	2.7 (mean)

team to improve the efficiency of processes at the two rolling mills.

Estimates indicate that implementing all ten projects could result in annual cost savings exceeding $11 million, corresponding to energy savings of ~1 million MMBtu and 38 million kWh annually. The annual projected savings for the ten projects are summarized in Table 1, below. After the table, brief discussions of the projects are presented.

1. *Convert the Lewisport mill to the continuous-casting process*: The assessment team recommended converting the Lewisport mill from direct-chilling casting to continuous casting integrated with the existing multistand hot mill. A 56% reduction in energy usage was estimated, allowing the mill to significantly reduce energy consumption, improving the sheet recovery from cast material and reduce working capital by eliminating the need for the scalping operation and ingot homogenization using soaking pits and tunnel furnaces.
2. *Upgrade melter/holder furnace at Newport mill*: The Newport melter/holder furnaces have dual purposes–receiving and holding molten metal delivered from the adjacent IMCO plant, and melting heavy gage and primary metal. Replacing three of four existing melters/holders with two furnaces in which melting and holding functions are optimized was recommended by the assessment team. The third melter/holder would be retained as a spare.
3. *Modify the production strategy to accommodate market demands*: The assessment team recommended that the mills minimize or eliminate less-than-capacity operations and optimize operations for which multiple processing devices exist, such as melting, melter/holder and holding furnaces, homogenization furnaces, annealing furnaces, and finishing processes.
4. *Implement best practices for melting operations*: The mills should define and implement best practices for melting operations, including minimizing time that furnace doors are open, improving burner firing, improving skimming, controlling exhaust gas temperature, improving maintenance of furnace sensors and controls, and improving staff communication.
5. *Improve melt stirring processes at Newport mill*: The assessment team recommended that the Newport mill upgrade its melt stirrer units to consume less energy and generate less dross.
6. *Upgrade soaking pits at Lewisport mill*: Recommendations to improve the efficiency of soaking furnaces included: (i) Replacing dry ceramic seals with more thermally efficient water seals; (ii) Upgrading the programmable logic controllers (PLCs) for the furnace; and

(iii) Using variable-speed recirculation fans and reverse air flow to mimic tunnel furnaces.

7. *Improve annealing operations*: Measures to improve the annealing processes included:
   - The Lewisport mill should evaluate methods employed at Newport for controlling annealing operations by using thermocouples on each coil interfaced with a PLC. The system permits focusing on the slowest heating coil while preventing overheating in other coils. This will increase energy savings and resulting in improved energy consumption.
   - Door seals should be improved to minimize heat losses at both mills.
   - The Lewisport mill should implement the Newport mill's practice of pulling coils from the furnaces at higher temperature to reducing annealing time and energy consumption.
8. *Optimize compressed air system*: Both mills should employ measures to upgrade and optimize their compressed air systems to improve system efficiency and maintenance practices, eliminate leaks, and substitute other energy sources for compressed air wherever possible.
9. *Use infrared imaging technology for process diagnostics*: The use of infrared imaging techniques is a powerful detecting heat losses and maintenance problems associated with thermal processes and electrical operations. These devices should be installed at both plants.
10. *Improve waste heat recovery*: Opportunities exist for improving waste heat recovery at both mills, primarily from melting, homogenization, annealing operations and paint line exhausts. Other specific opportunities should also be determined. Uses for recovered heat could include furnace air preheating, regenerative burners, steam generation, and hot water generation for applications within the mills or at nearby sites. The assessment team was unable to provide meaningful estimates of costs or savings associated with waste heat recovery because specific data were not available, although the team believed that these opportunities would be significant.

***Potential savings*:** total capital costs to implement all the projects were estimated to be ~$31 million. Implementing the projects with payback periods of 2 years or less was estimated to yield more than $2.4 million in annual cost savings.

Chemicals

3M Chemical Company: 3M Chemical Company performed a plant-wide energy-efficiency assessment at its Hutchinson, Minnesota, plant to identify opportunities for its plant's operations and utility requirements. This plant produces Scotch brand cellophane tape, and other consumer, office, industrial, and electrical supplies and tapes. The plant covers 1.3 million ft^2 consisting of two buildings (north and south buildings) and has more than 1600 employees. The plant spends about $7 million annually on energy-related expenses. Four separate implementation packages representing various combinations of energy-efficiency projects were identified from the assessment. The packages were developed separately because some of the individual efficiency improvement measures involving the thermal oxidizer units were mutually exclusive. One package (having the shortest payback period) was selected for implementation based on relative aggregate payback periods of the individual packages; this package included projects for chiller consolidation, and a heat recovery boiler for two of the plant's thermal oxidizers to produce low-pressure steam, which would offset the fuel requirements of plant steam production. Although the assessment team did not recommend immediate implementation of the packages involving combustion turbine-based cogeneration to destroy volatile organic compounds (VOCs), research on the technical feasibility of these measures was encouraged; this could help manage plant emissions while simultaneously improving energy efficiency. Energy efficiency improvements identified through the assessment would provide additional benefits by improving the plant's cooling efficiency and improving the reliability of chilled water production. The plant's productivity was also expected to improve because process loads will not need to be reduced if a chiller fails. The heat recovered for steam production will improve the thermal efficiency of the thermal oxidizer, reducing boiler fuel consumption and environmental emissions. Using a steam turbine to reduce steam pressure will offset the plant's retail electricity requirements.

The annual projected savings for the six projects are summarized in Table 2. After the table, brief discussions of the projects are presented.

Individual energy-efficiency measures identified during the plant-wide assessment and scheduled to be implemented as a package are described briefly below.

Cooling-chiller consolidation: The total capacity of the chilled water system is 7640 tons each for the north and south plants. By consolidating the chiller capacity of both plants (by interconnecting the individual chilled water distribution systems serving the plants), electric energy savings could be realized. The newer and more energy efficient chillers at the north plant could be used

Res–Solar

Table 2 Projects identified from 3-M Chemical Company's Plant-wide Assessment of its Hutchinson, MN, Facility

	Annual projected savings				
Project	**Electricity (kWh/yr)**	**Natural gas (MMBtu/yr)**	**No. 6 fuel oil (MMBtu/yr)**	**1st-Year avoided energy expense ($)**	**Projected capital cost ($)**
Chiller consolidation	1,552,750	N/A	N/A	87,420	292,546
Air Compressor cooling, North Plant	609,000	N/A	N/A	22,168	65,240
Air Compressor cooling, South Plant	393,750	N/A	N/A	34,287	170,775
Thermal oxidizer heat recovery boiler	N/A	38,093	172,557	772,191	913,275
Steam turbine	3,166,000	N/A	N/A	163,999	604,035
Relative humidity	N/A	695	3,145	14,200	0
Total	5,721,500	38,788	175,702	1,094,265	2,045,970

for base loads; these chillers could serve larger loads for longer periods of time, lowering operation costs. The chiller consolidation was estimated to provide an energy savings of more than 1.5 million kWh/yr.

Cooling-air compressor cooling north plant and south plants: The air compressors at both the north and south plants use chilled water for cooling; the nominal chilled water demand is 145 and 75 tons, respectively. The cooling towers of the chilled water system could be used as the primary cooling method for the air compressors–the potential energy savings equals the difference between chiller operation and operation circulating pumps. The electrical requirements were estimated to be reduced by 609,000 kWh/yr and 393,750 kWh/yr in the north and south plants, respectively, by implementing this measure.

Head recovery-thermal oxidizer heat recovery boiler: The 2L thermal oxidizer has a regenerative/recuperative cycle, thereby reducing the exhaust temperature and effectively eliminating the unit as a potential source for heat recovery. Units 1L and 3L, however, are suitable candidates for heat recovery applications. Two applications were considered–an oil-to-air heat exchanger to preheat supply air and makeup air, and a heat recovery boiler for low-pressure steam production. Because the payback period of the heat recovery boiler is considerably shorter than for the heat exchanger (1.2 years vs. 8.1 years), the heat recovery boiler was the recommended option. Using a reheat boiler as the heat recovery mechanism also mitigates the heat recovery limitations of an oil-to-air heat recovery system. With an oil-to-air system, the energy reduction annually is limited to the load of the specific air-handling units. Heat recovered from the thermal oxidizers to produce low-pressure steam has greater potential energy savings because the steam can serve loads throughout the plant. Annual steam production from the heat recovered from the two thermal oxidizers was estimated to be at least 164,000 million pounds annually; estimated savings are more than 210,000 MMBtu/yr due to the reduced need for natural gas and fuel oil.

Cogeneration-steam turbine: Steam is produced at a nominal pressure of 220 psig, but steam pressure is reduced to 125 psig and 15 psig for process loads and humidification, respectively. Using the pressure drop to drive a steam turbine and an electric generator (rather than reduction through a pressure reducing valve) could provide an offset in the retail electrical service requirement of the plant. Replacing the existing pressure-reducing valve with a steam generator provides an opportunity for cogeneration. More than 3 million kWh/yr in electricity were estimated to be saved by implementing this methodology.

Steam system-relative humidity Project: Steam provides part of the plant's relative humidity. From corporate guidelines, a relative humidity of 50% is necessary for the types of materials and processes used at this plant. Reanalysis has indicated that an acceptable range lies between 35 and 40%. Steam production can be reduced to correspond to this allowable range; no capital investment is necessary for this project.

***Potential savings*:** The plant was estimated to save 5.7 million kWh/yr in electricity and 214,499 million Btu/yr in natural gas and fuel oil by implementing the energy savings measures.

Forest Products

Blue Heron Paper Company: Blue Heron Paper Company and Pacific Simulation performed a model-based energy assessment to analyze effluent flow and heat load reduction, fresh water usage minimization, and process energy usage reduction at Blue Heron's paper mill in Oregon City, Oregon. Blue Heron was originally known as the Hawley Pulp and Paper Mill, which began operations in 1908. Times Mirror purchased the facility in 1948 (known as Publisher Paper Company). Smurfit Stone Container Corporation bought the mill in 1986, operating it for about 14 years. The mill was sold in 2000 to its employees and KPS to form the Blue Heron Paper Company. The mill produces ∼650 tons of paper daily, making newsprint and specialty paper on three paper machines. Approximately 60% of the fiber furnished to the paper mill is cycled fiber from old newsprint and magazines. Residual wood chips from sawmill waste are an additional primary source of fiber supplied to the thermomechanical pulping (TMP) lines, which convert them to pulp. Six natural, gas-fired utility boilers produce steam for the mill. Two lines of TMP pulping operate at Blue Heron. Each line is a primary and secondary refiner to separate wood chips into pulp fibers suitable for making paper. Rotating plates compress the stock, creating friction thus producing heat and steam. The steam is recovered in one of the lines and is converted into clean steam using a reboiler and thermocompressor.

The deinking pulping process involves pulping, debris screening, and flotation deinking. In pulping, the stock is mixed with steam-heated water to form pulp at a given temperature. Debris is removed from the pulp via screening and cleaning operations, after which the pulp is washed. The pulp slurry is refined and mixed with surfactants to float the ink particles. The slurry is cleaned and screened again to facilitate as much ink removal as possible. Water is filtered from the pulp and returned to the front of the deinking process for reuse.

Five processes were investigated during the energy-efficiency assessment: the deinking facility, three paper machines, the entire white water system, the TMP mill, and the steam system. Opportunities for effluent flow and heat load reduction, fresh water usage minimization, and process energy usage reduction were investigated by building full-mill mass and energy balances. A full-mill balance was required due to the complex system of water and steam reuse at the facility. Data were collected for the various processes to generate a steady-state model representing the base case. After establishing, the base case was a "summer" case and an "average" case were considered. The summer case, based on the highest recorded incoming water temperatures for the previous summer months, was used to determine the effluent heat loading conditions. The average case was used to calculate the overall annual economic benefits for the projects being evaluated.

After establishing the base cases, several possible projects were analyzed and quantified in simulations. Potential projects analyzed included the use of additional heat exchangers and/or cooling towers, modification of process conditions and process flows, and process equipment modifications. Fifteen different projects were evaluated during the plant-wide assessment; of those, 14 were considered potentially feasible (although some projects are considered to be alternatives to other projects). The projects listed in Table 3 exclude those considered to be alternatives to more economically attractive cases.

Individual energy-efficiency measures identified during the plant-wide assessment scheduled to be implemented as a package are described briefly below.

1. *Recycle vacuum pump sea water on paper machines No. 1 and No. 4; heat shower water*: Liquid ringed-seal vacuum pumps were currently installed on all three paper machines. The heat generated by creating the required vacuum is absorbed by the vacuum pump seal water, thereby providing steam that can be recycled and used as a heat source. Because the seal water temperature may be increased to the point where it flashes to steam under the high vacuum, it is possible to recycle up to 90% of the water. The seal water can then used to heat incoming filtered water for use in the paper machines, either by utilizing heat exchanger or by using a cooling tower, to dissipate excess heat for additional water reuse options. This requires recycle pumps, a heat exchanger or cooling tower for each paper machine, or a combination of components, depending on which machine is being considered for modification. Space availability is also a factor.
2. *Heat shower water for paper machines No.1 and No.4*: This activity involved recycling 75% of the vacuum pump seal water from the No.4 paper machine, and routing the flow through the No.1 paper machine to use as vacuum pump seal water. This water can then routed to the de-ink process water clarifier showers to reduce filtered water usage and to decrease the net amount of steam required in the paper mill. The recirculated vacuum pump seal water for each paper machine can be used to heat water in the paper mill. The recirculated vacuum pump seal water for each paper machine can be used to heat water required for the paper machine. The costs for this project activity result from the acquisition of heat exchangers and piping for the water. The estimated capital cost is $1.0 million, but it could be higher because of equipment logistics. In addition to the benefits in reduced relative heat load, and energy cost reduction, this processing alternative reduces the overall effluent flow by approximately 1.6 million gallons per day.
3. *Recover heat from Uhle box effluents on No. 4 paper machine*: Uhle boxes help remove water from the forming felt as it traverses from the forming section toward the dryer section. The Uhle

Res–Solar

Table 3 Blue Heron Assessment Results

Project description	Fuel savings (MMBtu/yr)	Electricity savings (kWh/yr)	Cost savings ($/yr)	Capital cost ($)	Payback period (yrs)
1. Recycle vacuum pump and water for paper machines No. 1 and No. 4; heat shower water	N/A	40	Included in total of project directly below	103,000	Included in total of project directly below
2. Heat shower water for paper machines No. 1 and No. 4	124,000	N/A	315,000	1,000,000	3.5
3. Recover heat from Uhle box effluents on No. 4 paper machine	32,000	N/A	100,000	110,000	1.1
4. Recover heat from deinking effluent to No. 1 paper machine	37,000	N/A	125,000	375,000	3.0
5. Reduce operating temperature in deinking plant pulpers	87,000	N/A	230,000	10,000	Immediate
6. Recover heat from vacuum pumps, Uhle boxes, and (TMP) wastewater; cool effluent with incoming filtered water	74,000	N/A	1,150,000	2,000,000	1.7
7. Heat shower water with reboiler steam and vacuum pump seal water; heat de-inking pulpers with TMP waste water	254,000	900	1,050,000	2,500,000	2.4
Total	608,000	940	2,970,000	6,098,000	2.1 (aggregate average)

box downstream and in the proximity of the steam box removes a combination of shower water and water from the sheet. The water's temperature is between 115 and 120°F. The type of heat recovery system is economical and technically feasible for the No. 4 paper machine. The capital cost is estimated at ~$110,000 with significant heat recovery, creating a project return on investment of 1.1 years. Heat recovery is more practical when applied to the Uhle box stream coming off the felt. The stream has the steam box on it and thus has the hottest Uhle box flow.

4. *Recover heat from de-ink effluent to No. 1 paper machine*: The de-ink plant's combined effluent stream at a temperature of ~120°F at a flow of 600 gal/min (gpm). These streams can be collected, passed through a new heat exchanger in the de-ink plant, and cooled with filtered water. The warm filtered water would displace steam that heats shower water on the paper machines. The heat exchanger itself could recover 9 MMBtu/hr at average conditions. The benefit in steam economy is only 3.5 MMBtu/hr due to some heat being lost through the effluent system, and the shower water flow being only 300 gpm. This stream is severely contaminated, so designing for minimal fouling risk increases the estimated capital expense to ~$375,000.
5. *Reduce operating temperature in de-inking plant pulpers*: The de-inking plant uses live steam to heat water in the pulper supply tank to a temperature of 125°F. The energy in the steam input varies from 9 MMBtu/hr in summer to 16 MMBtu/hr in winter; almost all of this heat is added to the mill effluent. The de-ink staff are considering operating at reduced temperatures; each 5°F decrease in temperature could save ~4 MMBtu/hr of steam use. Little capital expense is required for this modification.
6. *Recover heat from vacuum pumps, Uhle boxes, and TMP wastewater; cool effluent with incoming filtered water*: Options that could reduce energy consumption include combining recycling and cooling of vacuum seal water, heating paper machine shower water with TMP hot wastewater, recovering heat from Uhle boxes on the No. 3 and No. 4 paper machines, and cooling the effluent with incoming fresh water.
7. *Heat shower water with reboiler steam and vacuum pump seal water; heat de-ink pulpers with TMP wastewater*: The following operational changes would be required:
   - Preheating paper machine shower water with vacuum pump seal water on a partially closed cycle;
   - Heating paper machine shower water to operating temperature by direct injection of clean

Res–Solar

steam from the reboiler at 11 psig, thereby reducing power consumption in the steam compressors; and
- Heating de-ink pulpers by replacing some of the white water from the central wastewater from the TMP mill economizer.

This alternative requires less equipment than any other alternative considered that has comparable energy economy, due to much of the heat recovery is accomplished without heat exchangers. The effluent flow is reduced by 1.2 million gallons/day. The net steam consumption to heat stock and shower water drops to ~65 MMBtu/hr. A reduction in electricity consumption of ~900 kWh can be achieved.

Glass

Anchor Warner Robins: Plant-wide energy assessments were conducted at the Anchor Warner Robins, Georgia and Jacksonville, Florida plants identifying opportunities that can result in significant annual energy savings. Anchor Glass Container Corporation is the 3rd largest manufacturer of glass containers in the U.S. The Warner Robins facility has two furnaces and eight bottle-forming machines that produce more that 4 million bottles per day for approximately 360 days/yr. The facility contained 880,000 ft^2, comprised of 360,000 ft^2 for bottle production, 500,000 ft^2 for the finished-goods warehouse, 12,400 ft^2 for plant utilities, and 19,000 ft^2 for offices and miscellaneous space. Typical electric and gas loads are about 12.5 megawatts/day and 4 million ft^3/day. The Jacksonville plant was reduced in size during the past few years, consuming about half as much energy as the Warner Robins plant. Both facilities produce similar products. The primary materials used in manufacturing glass are sand, soda ash, limestone, cullet, and various corrugated packaging materials.

Anchor's furnaces are primarily gas-fired furnaces with electric booster heat. They melt more than 800 tons of glass/day. The furnaces are equipped with heat recovery regenerators that recover a portion of the waste energy from the 2800°F furnaces using a cycling air flow system. Fluctuations in the temperature differential between the heat recovery masses and the air streams during each 20-min cycle limit the effectiveness of the heat recovery process. Additional equipment that uses notable quantities of electricity include air compressors and vacuum pumps [7900 (hp) total, typically 6050 operating hp], cooling and furnace air blowers (3200 hp total), and lighting. The cooling water system, conveying and packing machinery, raw materials handling equipment, and a limited amount of space conditioning equipment also consume energy. Waste minimization and continued environmental compliance are goals for Anchor Glass. Environmental regulations pertaining to glass manufacturing include the disposal of checker slag, furnace residue removed during furnace rebuilds, furnace bricks containing chromium, and production waste. Additional regulations pertain to the discharge water used to clean machines, cooling water, dust produced by the batch mixing process, and air emissions from furnaces.

A systems approach was used to perform the plant-wide energy efficiency assessments. Opportunities were identified, evaluated, and prioritized for energy savings. Maintenance and operating procedures were also reviewed for their impact on energy efficiency. Three areas were considered in performing the energy assessments: inputs to plant processes, plant process efficiency, and process outputs including waste and heat products. Emphasis was placed on processes and systems identified to have the greatest energy savings potential:

- Cogeneration—installation of gas turbines with waste heat recovery systems;
- Waste heat recovery—recovery and reuse of waste heat from furnaces;
- Motor analysis—motor efficiency improvements;
- Lighting systems—cost-effective lighting improvements; and
- Variable speed drive (VSD) analysis—the installation of VSDs on selected process equipment, particularly blowers.

The results from the site-wide energy efficiency assessment are summarized in Table 4 (for both the Warner Robins and Jacksonville plants). The projections of energy and capital savings are preliminary, based upon the data collected and analyzed during the energy assessment; the data have not been validated through definitive engineering analysis and field testing.

Individual energy-efficiency measures identified during the plant-wide assessment scheduled to be implemented as a package are described briefly below.

Table 4 Potential energy savings summary

Action	Estimated energy savings/yr	Simple payback on capital required (yrs)
Increasing heat recovery	220,388 MMBtu	1.0
Compressed air	2,056,250 kWh	1.2
VSD cooling water pumps	524,600 kWh	1.8
VSD furnace air blowers	808,400 kWh	1.7
VSD machine cool blowers	560,720 kWh	1.8

Res–Solar

1. *Thermal Cycle Efficiency Improvement Options*: Heat is introduced to the glass furnace via direct natural gas over the glass melting tank, and electric booster heat arcing between electrodes immersed in the glass melt. Direct firing has lower cost fuels, but the over-melt firing process is much less efficient (50–60%) in delivering heat to the melt. The electric arc is nearly 100% efficient in delivering heat to the glass melt, but the cost of electricity is high relative to other fuels. To recover some of the heat lost through inefficiency of the direct firing process, the glass furnaces employ checker brick regenerators to capture and return some of the waste heat. The regenerators are characterized by their fluctuating effectiveness over the charge/discharge cycle; the regenerators are more effective at the beginning of the cycle because of greater air/regenerator temperature differences, but performance significantly decreases toward the end of the cycle. Further, significant heat is lost up the stack, providing an opportunity to improve process efficiency. Possible methods to improve the glass melting process efficiency include:
   - Generate steam from the stack gas waste to drive a turbine;
   - Recover stack gas waste heat into the incoming air through air-to-air or intermediary heat exchangers; and
   - Optimize charge/discharge cycle times to improve regenerator effectiveness.

   The assessment team considered recovery of the waste heat from the furnace stacks, to produce steam driving a turbine/generator and/or preheat incoming furnace combustion air. Their analysis indicated that the best way to recover the waste heat energy was by incorporating an air-to-air heat exchanger that transfers part of the heat between the air streams exiting and entering the regenerators.
2. *Compressed Air System Efficiency and Improvements*: Significant amounts of compressed air are used at two different pressures in container production processes at their facilities; high-pressure air (~100 psig) is used for operating typical air-powered controls and tools, and low-pressure air (~50 psig) is used in the bottle-blowing process machinery. A flow rate of ~16,000 and 11,000 ft^3/min of high-pressure air is used at the Warner Robins and Jacksonville plants, respectively. The energy assessment indicated that the system operates between 97 and 103 psi in the air compressor room at the Warner Robins plant; a refrigerated dryer rated at 11,000 cfm provides air-drying. The air supply in the plant supply header ranges from 82 to 85 psig. Air leaks were estimated for ~20% of the total air consumption. Efforts should be focused on reducing or eliminating the leaks to minimize air use. Opportunities identified to improve the efficiency of the high-pressure system included:
   - Remove system bottlenecks and add a higher-pressure storage system to reconfigure the compressor room to a lower supply pressure;
   - Provide advanced system controls to optimize compressor operators and energy efficiency; and
   - Reduce leaks in the system to lower flow requirements.

Adding the higher-pressure air receiver storage capacity will allow sufficient air storage so that the compressor can shut down for 10–20 minutes when unloaded, rather than cycling between loaded and unloaded modes.

Motor management and efficiency: The application and operation of electric motors were reviewed, resulting in the following recommendations:

- Economic guidelines should be in place so that initial cost does not drive repair vs. replacement decision-making;
- Larger motors that are sometimes underloaded for their application and are less efficient should be replaced with premium efficiency motors when a payback of 2.5 years or less is possible;
- An overall motor inventory and replacement plan should be developed; and
- Energy-efficient motors should be used for all new or replacement motors when expected annual operating time exceeds 4,000 hr.

Pump and blower VSD application: These facilities use pumps and blowers for various process functions, but three general applications were considered in the assessment: cooling water pumping, process cooling blowers, and furnace air blowers. These pumps currently run at constant speed with either no flow control or with valves for pumps and variable inlet vanes for blowers. Potential options to improve energy efficiency include:

- Use VSDs on the cooling tower water pumps to control pump speed and modulate tower water flow and pressure;
- Use VSDs on the glass furnace air and stack draft blowers to control blower speed as a means of modulating air flow; and
- Use VSDs on the glass forming and hot gas handling machinery air cooling blowers to control blower speed as a means of modulating air flow.

Plant lighting: Improvements in efficiency of plant lighting systems include the following options:

- Provide for expanded use of natural lighting
- Where convenient, convert the plant's outdoor lighting from utility lighting to wall packs or plant-supplied lighting
- Install motion sensors in equipment rooms, warehouse facilities, and other plant areas that are not frequently occupied
- Have lighting performance contractors conduct no-cost reviews of plant lighting to identify possible improvement opportunities.

Plant energy purchase optimization: The assessment team evaluated the potential for fuel substitution (including load management on the electric boost for real-time electricity prices and operational considerations). Break-even guidelines were recommended for consideration of energy purchase optimization and fuel substitution to permit better fuel management for these facilities.

Potential savings**:** Estimated total potential energy savings were 220,000 MMBtu/yr and approximately 4 million kWh/yr for electricity if all the project recommended were implemented. The associated capital required to achieve the fossil fuel savings was estimated to be $800,000, while the capital required to achieve the project electricity savings was estimated to be $250,000. Average simple payback periods calculated for the primary recommendations ranged from 1 to 2 years.

Metal Casting

AMCAST Industrial Corporation: A plant-wide energy assessment was performed to identify energy savings opportunities at AMCAST Industrial Corp.'s Wapakoneta, Ohio, manufacturing facility. The assessment highlighted process performance and its impact on overall energy and cost savings. The plant was founded in 1866, and is a major supplier of aluminum permanent-mold cast suspension components for the automotive industry; it also serves the construction industry and other industrial sectors. The plant employs nearly 390 people, and processes 20 to 25 million pounds of aluminum yearly.

Twelve separate projects were recommended to improve the efficiency of the plant's production processes and decrease the plant's energy consumption. The aggregate potential of these projects could save the plant $3.7 million, which would lead to a 3-month. payback period. Implementation of these twelve projects has the potential to reduce the plant's CO_2 emissions by more than 11 million pounds per year.

The plant performs heat treating and liquification of aluminum. The raw material is aluminum ingots. Natural gas-fired reverb furnaces melt the aluminum, which is transferred to hold furnaces adjacent to each low-pressure permanent mold machine via heated ladles. After casting, flash and scrap parts are sent back to the jet-melt furnace. Cast products are then solution and age-heat treated in special ovens, trimmed, inspected, and shipped to customers or sent to other locations for additional processing. Primary waste streams include aluminum dross, recyclable aluminum flash, de-burring material, metal shavings, and cooling wastes.

The facility's energy use, from transformers and switches to production lines, was monitored during the energy assessment. Data drawn from the operations were downloaded to data loggers, where it was evaluated utilizing software and documented. The information collected included:

- Thirty-day snapshot of peak kW demand;
- Peak daily kW demand (based on 30 minute readings);
- Metered values by phase (current, voltage, and power factor);
- Circuit-loading summary;
- Energy and power summaries; and
- Harmonic and disturbance analysis.

Opportunities to improve the reliability of the plant's motor systems, and cases where equipment should be added or replaced, were identified during the plant energy assessment. Energy data were used to calculate the expected savings and payback. The initial focus was on identifying and minimizing end-use loads. The distribution system was examined for inefficiencies and savings potential. After the end use and distribution systems were analyzed, the problem areas were examined for savings opportunities. In most cases, end-use and distribution savings directly influenced the recommendations for modifying the energy source.

The following recommendations were established from the plant-wide energy assessment:

- *Electrical systems*: Three-shift operations allowed the opportunity to move noncontinuous operations with large electrical demand to off-peak periods.
- *Lighting*: Levels were recorded and compared to recommended lighting levels. Options recommended included:
  - — Disconnecting obstructed lights or lights in overly lit areas;
  - — Making better utilization of task lighting and skylights; and
  - — Replacing T12 lamps and magnetic ballasts with T8 lamps and electronic ballasts in the offices.
- *Motor drive systems*: Motor repair and replacement policies were examined, and options recommended included:
  - — Use premium efficiency motors;
  - — Install variable speed drive (VSDs) on the casting machines; and

Res–Solar

— Replace belts on the motor drives with notched v-belts.

- *Compressed air system*: The compressed air system was examined for savings potential from using outside air, reducing header pressure, reducing pressure loss in the distribution system, optimal staging of compressors, and effective use of cooling air and water.
- *Processing heat*: Several options were identified:
  — Optimizing combustion efficiency of the ovens and furnaces;
  — Using waste heat and pre-heat combustion or space-heating air; and
  — Using solar walls to preheat process and space-heating air.

Once the data collection and analysis was complete, the assessment team developed twelve separate recommendations for projects. The team validated and supported the implementation of a number of these initiatives. The project recommendations resulting from the assessment are summarized below:

1. Switching to aluminum-titanate riser tubes to replace the Dense Fused Silica (DFS) material. The initiative results in reduced maintenance; less scrap, less downtime; and better product quality. It is estimated that savings in reduced scrap could exceed $1 million per year.
2. Using electric infrared heaters instead if existing gas torches to preheat permanent molds would further reduce scrap and save ~$850,000 annually.
3. Improving tooling design, repair and tool maintenance could lead to annual savings of ~$730,000.
4. Reducing scrap rate by addressing quality and scrap controls and increasing the number of quality control personnel to reduce rejects, could provide a projected net savings of $470,000 per year.
5. Using exhaust heat from reverbatory melting furnaces in heat treating furnaces, was estimated to provide an natural gas savings of $157,000.
6. Using exhaust heat from heat treating furnaces in aging ovens was estimated to provide an annual savings of $93,000.
7. Implementing other energy savings opportunities for reverbatory furnaces (such as capturing flue gas losses, radiation losses, and eliminating air leakage from furnace openings, etc.) was estimated to provide an energy savings of $64,000 annually.
8. Relocating jet melt furnaces and improve jet melt process flow by moving the furnaces closer to the jet melt- recycling unit would result in savings in energy, labor, and maintenance projected to total $89,000 per year.
9. Installing VSDs on casting machines was estimated to provide savings of $37,000 per year.
10. Using blowers to pressurize the cast hold furnaces was estimated to provide an annual energy savings of $11,000.
11. Using plant cooling tower water to cool the compressed air system instead of city water was estimated to provide $10,000 in savings of reduced municipal water fees.
12. Using exhaust heat from heat treating furnaces for scrap preheating (recapturing heat for reuse) was estimated to provide a savings of $3,300 annually.

Table 5 Wapakoneta estimated cost savings

Category	Estimated savings
Cost savings	$3.6 million/yr
Plant savings realized	$6.0 million/yr
Electrical energy savings	672,000 kWh/yr
Natural gas savings	9,000 MMBtu/yr

***Potential savings*:** Implementation of the above twelve projects was estimated to provide an annual energy savings of $3.5 million. Additionally, reductions in the utility costs could provide another $200,000 in savings. Since the total project costs are estimated to be ~$1.0 million the simple payback would be slightly more than 3.0 months. The estimated cost savings are summarized in Table 5 below. The payback periods for the twelve projects ranged from 0 to 29 months. Maximizing the plant's motor system energy efficiency would also reduce the amount of pollutants that the plant generated. Implementation of the twelve projects was estimated to reduce CO_2 emissions by 11,000,000 pounds.

Mining

Couer Rochester, Inc.: Coeur Rochester, Inc. mining operation in Rochester, Nevada, conducted a plant-wide energy efficiency assessment, in which five energy-savings opportunities were identified. Using energy cost tracking, process systems analysis, and characterization of the primary energy-consuming equipment to identify systems with the greatest energy savings potential, the assessment team identified five energy savings projects, focusing on improving the pumping systems, capturing waste heat, and upgrading lighting fixtures. It was estimated that 11 million kWh in annual electricity savings could be realized if the five projects were implemented, yielding an estimated cost savings of ~$813,000 per year. The required capital investment was estimated to be $260,000, resulting in an average simple payback of 4 months for all five projects.

Table 6 Cost Savings Estimated for Projects Identified during Coeur Rochester Plant-wide Energy Assessment

Project Description	Annual Electricity Savings (kWh/yr)	Implementation Cost ($)	Annual Cost Savings ($/yr)	Simple Payback Period (yrs)
Replace 1250-hp barren pump motor with a 600-hp motor	4,520,000	41,000	339,000	0.1
Replace 1250-hp phase 4 pregnant pump motor with a 750-hp motor	3,205,000	69,000	240,000	0.3
Install a regenerative system on downhill conveyor	1,920,000	115,000	144,000	0.8
Install adjustable-speed drive on 150-hp well pump No. 2	544,000	16,000	41,000	0.4
Install efficient lighting fixtures	648,000	19,000	49,000	0.4
Total	10,838,000	260,000	813,000	0.3 (average)

Coeur Rochester, Inc., is a subsidiary of Coeur d'Alene Mines Corporation. This facility opened in 1986, and is the company's largest producer and the largest primary silver mine in North America, producing 60,000 oz of gold and 6 million oz of silver annually. The ore body is blasted, and the ore is fed into a series of crushers to yield ore that is less than 3/8 inch in size. Conveyors transport this ore to the heap; a heap leaching process employing cyanide solution is used to capture the metals from the ore. The cyanide solution is pumped to the top of the heap and is distributed across the ore; the cyanide solution percolates down through the ore and captures the precious metals. The mine uses the Merrill-Crowe process to recover the metal from the cyanide solution. That process involves precipitation with zinc dust and is particularly applicable for gold ores with high silver content. Before gold is precipitated with zinc dust, all solids must be removed from the pregnant solution. The Merrill-Crowe plant has three principal stages: (i) deaeration to prevent gold from dissolving again; (ii) zinc dust precipitation; and (iii) precipitate filtration, in which zinc dust and gold are mixed with sulfuric acid to dissolve the zinc; the resulting mix is filtered, and the remaining solids are smelted into a bullion bar.

The assessment team tracked monthly energy costs, performed a process system analysis, and identified major energy consumers. The assessment involved the following activities:

- Performing a comprehensive review of product flow through the mine, from raw materials to final product, and establishing efficiency benchmarks;
- Developing a database of energy-consuming equipment;
- Identifying potential projects to reduce energy use and demand, and developing preliminary cost estimates for the projects;
- Prioritizing potential projects based on probability of implementation;
- Performing a detailed cost/benefit analysis for selected projects; and
- Developing implementation strategies for selected projects.

The above approach uncovered systems with the greatest potential for energy savings. The assessment team selected a number of projects for further study that involved major energy consumers, offered the greatest savings potential, and met a one-year simple payback criterion. Projects recommended for cost-savings potential included.

- Replace 1250-hp barren pump motor with a 600-hp motor;
- Replace 1250-hp phase 4 pregnant pump motor with a 750-hp motor;
- Install a regenerative system on downhill conveyor;
- Install adjustable-speed drive on 150-hp well pump No. 2; and
- Install efficient lighting fixtures.

Energy savings potentials for implementing these five projects are summarized in Table 6.

Individual energy-efficiency measures identified during the plant-wide assessment scheduled to be implemented as a package are described briefly below.

1. *Replace 1250-hp barren pump motor with a 600-hp motor*: The four-stage barren pump circulates 5500 gallons/minute (gpm) of cyanide solution to the top of the phase 2 and phase 4 heaps. Phase 2 requires 1300 gpm and phase 4 receives 4200 gpm. A 600-hp premium efficiency motor could replace the existing 1250-hp barren pump motor.
2. *Replace 1250-hp phase 4 pregnant pump motor with a 750-hp motor*: The phase 4 pregnant pump transfers metal-laden cyanide solution from the base of phase 4 to the processing plant. The head required is 500 ft, determined by calculating the pressure loss and elevation gain, using a 25 psi pressure drop through the clarifiers in the

processing plant. One bowl could therefore be eliminated from the existing pump configuration. A 750-hp premium efficiency motor could replace the existing 1250-hp pregnant pump motor. Additionally, an adjustable-speed drive could be installed to accommodate the change in flow requirements from phase 4 to the plant.

3. *Install a regenerative system on downhill conveyor*: A downhill conveyor transfers crushed ore to the base of phase 4 at an average rate of 1200 tons/hr. The energy created by the falling ore is currently dissipated as heat in the conveyor system's brakes; this energy can be recovered and used to generate electricity.
4. *Install adjustable-speed drive on 150-hp well pump No. 2*: Well pump No. 2 provides water to the processing plant for use in the plant and for use in suppressing road dust. Flow requirements fluctuate. A valve on the discharge line currently controls the flow rate. Installing an adjustable-speed drive on the 150-hp motor would allow better flow control and reduce the pump operating costs.
5. *Install efficient lighting fixtures*: Metal halide lighting fixtures are currently used throughout the shop and in warehouse spaces. Installing T-5 high-bay fixtures offer significant potential to provide cost savings and improve overall lighting performance.

***Potential savings*:** Implementation of the five projects was estimated to yield a cost savings of ~$813,000 per year. The required capital investment was estimated to be $260,000, resulting in an average simple payback of 4 mo for all five projects.

Petroleum

Valero Energy Corporation: Valero Energy Corporation performed a plant-wide energy assessment at its Houston, Texas refinery, involving an energy systems review to identify the primary natural gas and refinery fuel gas users, electricity, and steam-producing equipment, and cooling water systems, plus develop an energy optimization and management system. Valero Energy Corporation is headquartered in San Antonio, Texas, and has ~20,000 employees. Estimated annual revenues exceed $50 billion. The company owns and operates 14 refineries throughout the U.S., Canada, and the Caribbean. The refineries have a combined throughput capacity of more than 2 million barrels per day (approximately 10% of the total U.S. refining capacity). Valero bought the facility in Houston in 1997. This crude oil processing facility produces a wide range of petroleum products, such as gasoline, diesel fuel, kerosene, asphalt, jet fuel, sulfur, No. 2 and No. 6 fuel oil, and liquefied petroleum gas. It produces ~60,000 bbl/day of gasoline and 35,000 bbl/day of distillates. Processing capabilities primarily involve medium sour crudes and low-sulfur residual oils. The Houston facility is composed of several process units equipped with gas- or refinery gas-fired equipment. The units process raw crude oil into various finished products by means of distillation and chemical reactions. The units include the following:

- *Crude complex* consisting of atmospheric and vacuum distillation towers with two lightends removal towers;
- *Fluidized catalytic cracking complex* that thermally cracks heavy gas oil material into lighter products (such as gasoline and diesel).
- Hydrotreating/reforming complex that processes kerosene, diesel, and naphtha in four desulfurization units; the desulfurized naphtha is reformed in a fixed catalyst bed reformer to produce a high-octane gasoline blend component;
- Sulfur removal complex recovering sulfur gas streams in an amine contacting/regeneration unit and a sulfur plant before using the gas streams as refinery fuel gas; and
- Alkylation complex that produces a high-octane gasoline blending component using a sulfuric acid alkylation unit; an isooctane unit converts butane streams to isooctane for gasoline blending.

Refinery fuel gas is used in process heating and steam production. Four fixed boilers and waste heat recovery equipment (two cogeneration units) produce steam. Each process unit requires cooling water that is supplied by six refinery cooling water towers to remove low-level heat. The refinery's three flare systems vent and burn light gases produced by the process equipment. The facility has access to three water sources: the Coastal Water Authority (CWA) supply, City of Houston municipal water, and well water.

The energy assessment addressed ways to reduce water use and environmental emissions. Fourteen projects were identified for potential implementation at the facility. If all the projects identified were implemented, the potential energy savings to the Houston refinery was estimated to be 1.3 million MMBtu (fuel) and more than 5 million kWh (electricity). Total annual cost savings were estimated to be ~$5 million. The plant-wide assessment consisted of two primary activities:

- *Energy systems review*: The assessment team reviewed the primary energy systems (natural gas, refinery fuel gas, electricity production, steam, and cooling water). Major natural gas and refinery fuel gas users, electricity- and steam-producing equipment, and cooling water systems were identified. Energy-use data was collected to identify major energy-consuming equipment and processes.
- *Energy optimization and management system development*: Assessment of data and energy system information were gathered to develop a computer model of the primary refinery processes and the energy

production and distribution systems. The model can be used to determine the most efficient loading of individual pieces of equipment

Energy savings potentials for implementing these fourteen projects are summarized in Table 7 below.

Individual energy-efficiency measures identified during the plant-wide assessment scheduled to be implemented as a package are described briefly below.

— *Upgrade steam system insulation*: Using a computer model, the assessment team identified several uninsulated or underinsulated steam system lines and equipment; the upgrading the steam system insulation should be implemented to reduce energy losses.
— *Install new air compressor*: Two rented diesel air compressors with a new steam-driven compressor would result in saving ∼108 gallons/day of diesel fuel and $160,000/yr in rental fees.
— *Install CO control system*: Two Bambeck CO control systems on the crude unit's atmospheric and vacuum heaters were recommended by the assessment team. By reducing the excess oxygen from 3.5% to 1.5%, energy savings will result. Nitrous oxide (NO_2) and CO emissions would be lower due to the lower flame temperatures. The Bambeck CO control system uses a CO analyzer on the heater flue gas, a stack damper control system, heater draft instrumentation, and computer software to minimize excess oxygen.

Table 7 Estimated Project Cost Savings for Valero's Houston Refinery

Project Description	Cost Savings ($/yr)	Capital Cost ($)	Pay-back Period (yr)	Fuel Savings (MMBtu/yr)	Electricity Savings (kWh/yr)
1. Upgrade steam system insulation	168,100	100,000	0.6	51,906	N/A
2. Install new air compressor	200,000	400,000	2.0	5,125	N/A
3. Install CO control system	247,700	401,000	1.6	77,406	N/A
4. Install cooling tower automatic blowdown control system	130,000	240,000	1.8	N/A	N/A
5. Incorporate an energy optimization and management system	2,200,000	270,000	0.1	687,500	N/A
6. Inspect, repair, and maintain steam traps	656,000	100,000	0.2	205,000	N/A
7. Install boiler automatic blowdown control system	100,000	180,000	1.8	N/A	N/A
8. Replace boiler house boilers	161,000	3,000,000	18.6	50,313	N/A
9. Install flare gas recovery compressor	420,000	1,000,000	2.4	131,250	N/A
10. Install evaporative coolers on cogeneration units	174,700	249,000	1.4	−17,035	4,987,000
11. Clean the CWA water supply line	13,000	6,000	0.5	N/A	N/A
12. Assign an energy coordinator	200,000	130,000	0.7	N/A	N/A
13. Turn off outdoor lighting during daylight hours	6,000	N/A	0	N/A	175,000
14. Install liquid ring vacuum compressor	306,000	450,000	1.5	95,625	N/A
Total	4,980,500	6,526,000	1.3 (mean)	1,287,090	5,162,000

— *Install cooling tower automatic blowdown control system*: Instrumentation and controls on four cooling towers to automatically control each tower's conductivity (which provides a measure of the solids in the circulating water system) were recommended. High conductivity reduces makeup water costs, chemical use for water treatment, and downstream processing costs, but it increases cooling water system fouling and corrosion while reducing heat transfer. Low conductivity has the opposite effects. Maintaining the conductivity within a suitable range is important for overall performance and cost of the cooling water system. Makeup water and chemical costs were estimated to be reduced by $130,000.

— *Incorporate an energy optimization and management system*: An energy optimization and management system (EOMS) should be developed to use in assessing, implementing, and tracking process unit energy changes and new project implementation effects on the refinery. Implementing EOMS was estimated to reduce annual energy use by 2%–6%. The simulation and optimization software optimizes the purchase, supply, and use of fuel, steam, and power at the refinery, based on process unit energy demands and system constraints caused by equipment or environmental regulations. The software analyzes conditions such as supply contract variability, alternative fuel options, optimum loading of steam boiler equipment, motor vs. turbine driver decisions, and importing vs. exporting of steam, fuel, and power. The software performs the following functions for the refinery:
  - Facilitates optimal planning of utility equipment;
  - Assists in optimal operation of the utility plant and associated equipment;
  - Provides real-time information on site-wide energy performance, utility costs, and revenue; and
  - Provides real-time information for use in prioritizing maintenance tasks;

— *Inspect, repair, and maintain steam traps*: Leaking or failed steam traps cause significant energy losses and create process problems. Leaking or failed steam traps were recommended to be repaired and implementation of a maintenance program.

— *Install boiler automatic blowdown control system*: Installing instrumentation and controls on three steam boilers was recommended for automatically controlling boiler blowdown conductivity. High conductivity reduces boiler feedwater expenses and chemical use, but it increases boiler system fouling and corrosion, while reducing heat transfer. Low conductivity has the opposite effects. Controlling conductivity within a suitable range is important to overall boiler system operation and costs. Boiler feedwater and chemical feed were estimated to be reduced by ∼$100,000 annually.

— *Replace boiler house boilers*: Replacing the four 1940's-era 50,000 lbs/hr induced draft boilers with a new 200,000 lb/hr, boiler was recommended, to improve operating efficiency and reduce fuel gas usage. Additionally, NO_x and other emissions would be reduced, providing additional economic and environmental benefits.

— *Install flare gas recovery compressor*: Installing a compressor and associated equipment was recommended in order to recover flare gases from three existing flare systems. Recovered gas can be amine-treated to remove hydrogen sulfide and routed to the refinery fuel gas balance drum, thereby reducing natural gas makeup to the gas balance drum (except when there is excess recovered fuel gas because one or both off-refinery fuel gas purchasers are shut down).

— *Install evaporative coolers on cogeneration units*: Installing intake air coolers was recommended to increase the cogeneration of electric power or to sell to the local power grid. The cogeneration units will probably be replaced due to their aging components, inefficiency, and NO_x emissions.

— *Clean the CWA water supply line*: Cleaning a raw water supply line was recommended to reduce pressure drop and increase flow. This program has already been implemented; the system now supplies the added water needed to meet refinery demands without having to bring in more expensive municipal water.

— *Assign an energy coordinator*: Prior to conducting the plant-wide assessment, no one was responsible for monitoring refinery energy use and developing energy projects. A full-time energy coordinator and support personnel were recommended. This team would be responsible for such activities as ensuring that steam trap surveys are conducted, developing energy use simulation models, reviewing flare system losses, developing energy projects, and monitoring energy use.

— *Turn off outdoor lighting during daylight hours*: Lighting was being left on unnecessarily during the day. Operations personnel were reminded to turn off these lights when they were not needed. The team also found that most areas in the refinery are equipped with photocells. Several photocell switches had become dirty or had accidentally been painted over, so they were in need of repair or replacement.

— *Install liquid ring vacuum compressor*: Installation of a liquid ring vacuum compressor on the vacuum tower overhead system was recommended to replace the third-stage vacuum steam jets and water cooler. This project was estimated to reduce steam use at 285 psi by ∼6,000 lbs/hr. Additional energy savings should result from improved suction pressure on a downstream compressor.

***Potential savings*:** The projects recommended were estimated to save Valero Energy Corporation about

$5 million annually, while reducing environmental emissions.

Steel

Jernberg: To streamline its manufacturing processes at its Chicago, Illinois, facility, Jernberg chose to convert from its traditional batch methods to lean manufacturing. Jernberg was founded in 1937; it was the country's first independent press forging company. The company's forging facility located in south Chicago, produces 170 million lbs of forged parts annually. Jernberg produces a wide variety of gears, yokes, hubs, and other parts for the automobile and motorcycle industries. A plant-wide energy efficiency assessment was conducted to address energy-intensive processes requiring change. Efficiencies of the primary support systems (e.g., compressed air, cooling water, etc.) were also investigated. A process simulation model was used to evaluate the impacts of converting existing batch production to a lean manufacturing operation; it employed a systems approach to identify trends and energy use distribution for plant equipment. Seven projects were identified that were estimated to save Jernberg more than 64,000 MMBtu/yr in fuel and more than 6 million kWh/yr in electricity.

The primary raw material used in forging operations is steel bar. The bar stock is often preheated with a natural gas-fired burner to drive off moisture and to facilitate shearing. The heated bar is fed into one of seven shear presses or saws, where it is cut into billets of 6–15 in. lengths. The billets are deposited into metal boxes as they leave the shear press. The billets are transported by forklift to one of ten forging lines, where the billets are manually fed (charged) into a conveyor that feeds the billets into a pass-through induction furnace. The heated billet is ejected from the furnace discharge and drops down a feed chute that serves the forging press. At the bottom of the chute, a worker lifts the billet from the chute and places it on the bottom die. The forging ram is activated and compresses the billet. The pressed billet is lifted and placed on the second stage die where it is rammed into its final shape. Once the billet has been forged, the part is ejected from the press and conveyed to a trim press where flash is removed from the hot part.

About half of the parts that are produced are also heat-treated, which is performed in one of five batch-style heat treat furnaces. Several heat treat methods are used, including water-quench, oil-quench, normalizing, and annealing. Many parts undergo finishing operations such as shot blasting, drilling, grinding, magnaflux testing, dimensional testing, and ultrasonic testing.

The lean manufacturing methodology optimizes equipment utilization, maintenance programs, information flow through the plant, and workplace organization. Lean manufacturing requires plant personnel to assess ways to eliminate waste, including work-in-progress inventory, unnecessary movement of parts throughout the plant, overproduction, machine downtime, etc. Primary benefits of lean manufacturing include more efficient overall utilization of the plant, reduced inventory (and associated carrying costs), improved product quality, and improved equipment reliability. Jernberg traditionally produced forged parts using batch processes, which can length a production cycle time by 3 to 5 times or longer. The energy assessment team performed a process simulation to evaluate the impacts of converting existing batch production to a lean manufacturing operation. Heat-treating was identified as the primary bottleneck in the process. Using controlled cooling instead of batch heat treating largely eliminates that problem as well as provide significant energy savings. A systems approach was used to evaluate the plant's energy consumption. Historical data over a two-year period was evaluated to identify trends and energy use distribution for plant equipment; this allowed identification of systems to be targeted for further evaluation. Once systems and equipment were identified and targeted, recommendations were developed that were both technically and economically feasible. These included recommendation that supported conversion to lean manufacturing. The efficiencies of the primary support system were also evaluated, such as for compressed air and the cooling water loop.

The assessment team identified seven projects during the energy efficiency assessment. If all the projects were implemented, Jernberg estimated that it could save more than 64,000 MMBtu/yr in fuel and more than 6 million kWh/yr in electricity. Total annual cost savings were estimated to be ~$791,000. Total implementation costs were estimated to be ~$2,000,000 for all the projects. The specific projects recommended are listed below:

- Repair recuperator;
- Recover waste heat from cooling tower loop;
- Eliminate billet reheats;
- Install air compressor controls;
- Convert heat treat processes to controlled cooling;
- Replace air compressors; and
- Reduce forge press downtime.

Individual energy-efficiency measures identified during the plant-wide assessment scheduled to be implemented are described briefly below.

Repair recuperator: The existing recuperator on heat treat furnace No. 5 is not being used; it was taken out of service during a previous furnace rebuild operation. As a result, air at 1400°F is exiting the stack during furnace operation. By repairing the recuperator and returning it back to service, the stack exit temperature could be reduced to 400°F, and the recovered heat could be used to preheat the combustion air in the furnace, saving an estimated 1812 MMBtu/yr. The implementation cost was estimated to be ~$25,000, assuming that the recuperator would need to be repaired before returning it back to service. This project was estimated to provide an annual cost savings of $9,400; the resulting simple payback period would be ~2.7 yr.

Recover waste heat from cooling tower loop: Fourteen cooling towers provide process cooling for the induction heating furnaces, air compressors, and hydraulic systems on each forging line. Jernberg uses ~12,000 MMBtu of natural gas for space heating annually. Most areas of the plant require space heating (except for the forging area, which relies on waste heat from the process to warm the area during the winter). Of the many cooling loops used in the plant, only the loop serving the north air compressors is used consistently enough to provide a steady source of heat for the plant. This loop rejects 3650 MMBtu of heat from two Ingersoll-Rand compressors. Using this waste heat to warm the final processing department and displace the existing gas-fired unit heaters that are only 80% efficient would save 4,652 MMBtu a year (estimated). Annual cost savings are estimated to be $23,700 per year. The estimated cost to implement this project is $25,000, resulting in a simple payback period of 1.1 years.

Eliminate billet reheats: The induction heaters operate on an open-loop control basis to take room-temperature billets and heat them to 2300°F prior to delivering them to the forging process. When a downstream process goes down, the heated billet is diverted to totes where it is allowed to cool to room temperature before being reheated. The furnace has no automatic controls, so an operator must shut down the charging system that feeds the billet stock into the furnace to reduce waste heating of billets when the line goes down. A closed-loop has been proposed for each line to allow the induction furnaces to react to downstream variations and speed up or slow-down as needed to match production requirements more closely. This approach would tie in with the lean manufacturing process, as it would allow all processes in the forging cell to balance. Billet heating is by far the most energy-intensive step in the manufacturing process. Using a closed loop approach would eliminate the reheating step, providing an estimated energy savings of 4.1 million kWh/yr of electricity, with an annual cost savings of ~$247,700. The cost to retrofit all ten lines was estimated to be $500,000 (which includes the installation of controls and queue sensors as well as programming logic to operate the furnaces) providing a simple payback period of ~2 years.

Install air compressor controls: At the plant, Ingersoll-Rand compressors are base loaded while Fuller Rotary vane compressors operate partially loaded in throttle modulation. Throttle modulation is an inefficient mode of control. Plant personnel turn compressors on and off as needed. This generally causes more compressors to be on than are really needed, particularly during shift changes and when process lines are shut down early. An integrated control system would minimize the human factor from compressor control, and ensure that the vane compressors are base loaded when online, using the reciprocating compressors as swing machines. The electronic controls allow the compressors to maintain pressure at ±1 psig, allowing the compressor setpoint to be reduced safely within 2 psig of the plant's minimum pressure requirement. The combined energy savings associated with pressure reduction and capacity control was estimated to be 1.8 million kWh annually with a cost savings of ~76,900/yr. Using an implementation cost of $270,000, the simple payback period was estimated to be 3.5 years.

Convert heat treat processes to controlled cooling: Forged parts are allowed to cool a minimum of 24–48 hours; they are then reheated to 1500°F in the heat treat furnaces. Controlled cooling takes the hot forged parts and carefully cools then in a closed chamber. The cooling profile makes certain that the parts' microstructure is properly developed to ensure proper hardness, which is accomplished by using the residual heat from the forge process. A type of controlled cooling is currently used in the plant by utilizing vanadium alloyed steels. Implementing controlled cooling on other parts was estimated to save $7700 MMBtu annually by utilizing the process heat that is currently wastes for heat treating of parts. This system would readily fit in with the lean manufacturing approach due to the forged parts being fed via conveyor directly through the cooling chamber. Estimated costs to implement this project was $1.1 million, which includes $600,00 to model the cooling profiles for the parts, plus $500,000 to construct two chambers that provide the cooling process. An annual natural gas savings was estimated to be $352,000 yielding a simple payback period of 3.2 years. This project was deemed to be highly important, since it would eliminate a production bottleneck in the heat treat department and would increase the plant's capability such that raw materials could be converted into finished product within one work day.

Replace air compressors: Four existing Fuller air compressors are at the end of their service life. The machines are two-stage rotary vane compressors. Due to their age and wear, maintenance cost on these compressors is high (~$10,000/yr for materials only) and the machines' capacity has degraded because of increased blowby between the vanes and the chamber wall. These compressors are water-cooled; there are additional water costs associated with cooling tower evaporation and drift. Replacing these aged compressors with new, two-stage rotary screw compressors was recommended to increase compressor specific power from 4.3 cfm/bhp to 5.3 cfm/bhp at the application conditions. This improvement in efficiency was estimated to save 143,100 kWh/year with a cost savings of $74,000/yr. The cost of four new compressors was estimated to be $280,000, resulting in a simple payback period of 3.8 years.

Reduce forge press downtime: When dies are changed in the forging process, new dies are heated using a natural gas torch that is sandwiched between the die cavities for about 20 minutes. When die changes are required during production hours, this time represents lost productivity. If an infrared die heating station is installed, the same die heating process could be accomplished in 4 to 5 minutes, enabling the die to be changed and return the press back in to production more quickly. Some energy savings would be realized by eliminating the torch. The primary savings will occur as a result of less downtime, thereby reducing billet heater losses. This savings was estimated to be ~$7,200 annualy. The cost of an infrared die heat station was estimated to be $5,000, yielding a simple payback period of 0.7 yr.

***Potential savings*:** With implementation of the seven projects, Jernberg estimated that it could save more than 64,000 MMBtu/yr in fuel and more than 6 million kWh/yr in electricity. Total annual cost savings were estimated to be ~$791,000. Total implementation costs were estimated to be ~$2,000,000 for all the projects.

Geothermal Plant

Alaska's first geothermal plant was dedicated at Chena Hot Springs, Alaska [see website: www1.eere.energy.gov/news/progress_alerts/progress_alert.asp?aid=183]. This plant is the first to use a new technology making electricity generation possible at lower temperatures. The plant makes geothermal power plants feasible in many more locations than today's high temperature technologies. The technology is a pure cycle organic Rankine device to convert waste heat and liquid streams in to power to increase system efficiency of distributed generation devices. At Chena Hot Springs, the system will offset 160,000 gal of diesel fuel annually, representing a cost savings of ~$384,000.

Res—Solar

BIBLIOGRAPHY

1. U.S. Department of Energy, Energy Efficiency and Renewable Energy, U.S. DOE website: http://www1.eere.energy.gov/industry/bestpractices/printable_versions/case_studies.html (accessed January 2007).
2. U.S. Department of Energy, Energy Efficiency and Renewable Energy, U.S. DOE website http://www1.eere.energy.gov/news/progress_alerts/progress_alert.asp?aid=183 (accessed January 2007).

Savings and Optimization: Chemical Process Industry

Jeffery P. Perl
Chicago Chem Consultants Corp., Chicago, Illinois, U.S.A.

Robert W. Peters
Department of Civil, Construction & Environmental Engineering, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.

Abstract
For the chemical process industry (CPI) to remain competitive in the global environment, both environment, safety and occupational health (ESOH) and process design must be integrated and simultaneously optimized. A practice-based approach to chemical process optimization is outlined emphasizing these concerns. Raw-material usage and energy consumption must be balanced. Mass and energy balances are coupled in order to use each in the most efficient manner possible. Process energy is optimized by consideration of process heat integration, low- and high-temperature sources, mechanical systems (e.g., microturbine power generation, pumping, compressed air, hydraulic systems, pneumatic transport, and chillers), electrical systems, water, utilities, waste-to-energy opportunities, and facility energy integration. These compounds can be applied to CPI unit operations, such as distillation, liquid extraction, drying, freeze crystallization and drying, crystallization, filtration, chemical reactors, etc. Several case studies integrating mass and energy optimization are reported related to process energy optimization. A checklist is presented for optimizing process energy systems.

ENERGY UTILIZATION AND SAVINGS IN THE PROCESS INDUSTRY

Introduction

The business of the chemical process industry (CPI) is turning raw materials into finished products, usually on a large scale. The CPI includes a wide variety of manufacturing sectors including, food, pharmaceutical, plastic, oil refining, and building materials to name only a few. All have as a common requirement a need to balance raw material usage with energy consumption as well as Occupational Health and Safety Administration (OSHA) and Environmental Protection Agency (EPA) regulatory compliance. Product quality assurance requirements must also be met. It will be helpful to review some of the regulations that affect process energy considerations.

Keywords: Chemical process industry; Environment, Safety and Occupational Health (ESOH); Material and energy balance; Energy optimization; Process heat integration; Heating, ventilation, and air conditioning (HVAC); Variable-frequency drive (VED); Unit operations; Combined heat and power; Cost; savings; Electricity; Natural gas; reaction engineering; Building energy management; RCRA; Occupational Health and Safety Administration (OSHA); Environmental Protection Agency (EPA); Process troubleshooting; Low-quality heat; High quality heat; Turbine; Power.

U.S. EPA Regulations

Resource Conservation and Recovery Act of 1976 (RCRA), United States Environmental Protection Agency (EPA) RCRA was crafted exactly during the time of the first (1973) energy crisis in the United States. This happened when oil-exporting countries embargoed and severely curtailed oil exports and raised prices. As a result, the U.S. Department of Energy (DOE) began an extensive program to promote alternative energy sources including novel methods to recover more of the 60%–80% of crude oil left behind after conventional drilling and pumping operations. In light of these contemporary activities, RCRA was written with energy prominently mentioned in the opening paragraphs.

Selected Excerpts from the original RCRA statute:

> Section 1002 (d) ENERGY—The Congress finds that with respect to energy:
>
> 1. Solid waste represents a potential source of solid fuel, oil or gas that can be converted into energy;
> 2. The need exists to develop alternative energy sources for public and private consumption in order to reduce our dependence on such sources as petroleum products, natural gas, nuclear and hydroelectric generation; and
> 3. Technology exists to produce usable energy from solid waste.

Objectives and National Policy

Section 1003 (a) OBJECTIVES—to promote protection of health and environment and to conserve valuable material and ENERGY resources…

We can see elements of this philosophy on U.S. EPA boiler industrial fuel (BIF) regulations allowing companies to burn used oil onsite to recover energy value, while maintaining clean air compliance.

US Department of Energy Program

The federal government has an excellent program to assist both commercial and residential end users of energy. Information is available for US DOE as well as US EPA Energy Star program on the web. Much information has been developed including efficiency of equipment and appliances as well as actual energy project cost and effectiveness study results. We have presented this information at the end of this article and our following article.

COMBINED ENERGY AND ENVIRONMENT, SAFETY AND OCCUPATIONAL HEALTH (ESOH) CONSIDERATIONS

As we have seen from the very inception of US EPA, both energy and environment have been linked. For the design engineer, it is imperative that both ESOH and process design be simultaneously optimized. Sustainability captures a global concept and includes such elements as raw material and prime energy resource availability into the planned project future. Health and Safety also are of primary concern up front in any process design. We lump all such considerations together as ESOH.

When making process changes to reduce energy consumption, take care to avoid negative impact on the product. This might entail re-testing for intermediate and end point specification. International Standards Organization (ISO) considerations may also be involved regarding total quality assurance. OSHA management of change (MOC) regulations will also come into play to ensure worker safety is maintained or improved, and always in compliance. The EPA risk management program (RMP) requirements also will come into play. The details of these OSHA and EPA programs are beyond the scope of this article, but their importance is not. The law of unintended consequences is be best managed by understanding the interplay of elements involved in production.

MATERIAL AND ENERGY BALANCE… WHERE IT ALL BEGINS

The very first course of study for chemical engineering students is Material and Energy Balances (MEB). When designing a chemical process, a form of "double entry accounting", evolved regarding materials processed and energy consumed. These two elements are always coupled and provide the designer with a convenient method to both design the process as well as double check and optimize the use of both materials and energy.

A simple example is that of boiling water. Here approximately 1000 Btu is required to evaporate a of water. A process designed to distill water, therefore, must account for both providing this energy (source) and dissipating this heat (sink). The heat going into a process minus the heat leaving must be equal to that consumed by reaction requirements or non-steady state energy buildup. So a source such as a boiler must be selected to supply needed heat and a sink such as a condenser must be selected to remove the heat. This calculation can also be used to provide extremely close digital control of the process through coupled feed forward and feedback control loops, with the calculated energy needs (feed-forward) being fine-tuned by the feedback measurement of actual conditions, as tempered by adaptive controls. Detailed knowledge of the process MEB information also identifies energy inefficiencies and can become the basis for improvement in operation.

The material balance alone can provide useful information about the health of the process. Material entering the process minus that leaving must be equal. Differences can arise from non-steady state conditions such as accumulation in a tank. Any unaccounted difference might be considered lost material and may serve as the basis for environmental permit violation as well as bottom line process economic loss.

An energy balance alone can also be useful. For example, if you cannot account for 20,000 Btu of supplied heat, it may indicate a loss of 20 lb of water or poor flow or temperature measurement. A typical steam energy balance problem might involve inadequate or missing condensate return, poor check valves, traps, etc. Similar problems come up in chemical processes such as plugged or leaking heat exchanger tubes, partially shorted pump motor windings, broken mixing blades, poorly distributed or destroyed mass transfer packing, deactivated catalysts, and buildup of harmful byproducts (such as scale and chemicals) that reduce reactor efficiency or selectivity.

So it becomes apparent that while independent material and energy balances can be useful, they really shine when ever coupled as a double check of process or utility health.

If our simple water-boiling example became a solvent distillation instead the unaccounted energy could be tied to solvent loss with attendant environmental implications or worker health impacts, so the MEB is extremely important, not just in design, but as a barometer of process health and efficiency as well. The attempt to address environmental losses through waste minimization must therefore be accompanied by equal consideration to energy optimization. In some instances, energy will be

Res–Solar

the primary consideration (leading indicator), and in others, material will be.

We will focus more on energy as a leading indicator. An unexpected benefit of process energy review and optimization provides the designer an opportunity to reevaluate the entire system.

PROCESS ENERGY OPTIMIZATION

The overall process of turning raw materials into finished products is energy intensive. Processes consume mechanical and electrical energy as well as water, natural gas, oil and other combustible commodities. Waste byproducts are also produced that must be either treated or disposed of, often at the loss of their energy contents.

Process Heat Integration

Chemical manufacturing is comprised of physical unit operations to separate and concentrate chemicals without changing them, while reactors change raw chemicals into desired products such as, plastics, fuels, pharmaceuticals and other products. Unit operations such as distillation and drying consume large amounts of energy and cooling water. Reactors commonly produce heat (some are endothermic, or energy consuming) and use cooling water to maintain temperature. Mechanical, electrical, power, and heat engine utilities exist to meet process needs. In addition, facilities need heat in winter and cooling in summer, operations that both consume energy. Through careful management and design, the various energy sinks are matched to sources to minimize outside requirements of electricity, gas and oil.

Low- and High-Temperature Sources: Low versus High-Quality Energy

Thermodynamic engines require large temperature differences to produce power, and they still have relatively low energy conversion efficiency. The key to high overall thermal efficiency is to couple a steam-driven electric turbine to a process that requires low-quality heat. With the advent of the microturbine and utility electricity buyback regulations, individual facilities can now consider power generation. When you couple in transmission line loss of up to 30%, it should make on-site generation projects increasingly more attractive. Those that have small process heat requirements might consider locating near or attracting onsite manufacturers that have high heat demand. The same arguments hold for plant chillers and facility HVAC air conditioning system combinations.

Boiler heat exhaust can be used for instance to create high-pressure electrical turbine steam. However, many streams requiring cooling, such as unit operations and reactors and do not meet these differential temperature requirements. In these instances, we look for streams requiring preheating to match against streams requiring cooling.

Facility HVAC equipment, such as chillers and boilers can often exhaust/supply low quality energy to chemical processes at a very high efficiency. When designing a chemical process, engineers typically place collection vessels before and after to take up surges. In fact, the concept is similar to an electrical surge protector. Surge capacity must be engineered if a facility generates electricity for use on- and off-site, or when integrating facility heat and cooling loads with process stream energy demands.

ELECTRICAL, MECHANICAL, AND OTHER UTILITY DESCRIPTIVE CHECKLIST

It is difficult sometimes to separate these elements from the process itself, but it is helpful, as most plants are divided this way. Let us examine some common areas of energy interest. These topics are covered elsewhere in this encyclopedia, but they deserve mention here to identify the method and manner that they interact with and serve the process.

Mechanical

Microturbine Power Generation

Recent advances in electric generation technology have brought this down to the personal home size. In chemical processes, one can generate power from waste heat to steam. The key to success lies in either using generated power for your plant, selling it to the power company or both. Remember to factor in low-quality waste heat that may exit the turbine. This can be captured as preheat for the process or to preheat boilers or hot water heaters or even as seasonal facility heating.

Pumping

Virtually no chemical process is without some sort of pump. Whether liquid, solid, or gaseous, a variable-speed pump or blower versus a fixed-speed pump and valve arrangement can reduce or eliminate expansion/throttling losses. Here the ubiquitous variable-frequency drive (VFD) can take the place of energy consuming throttle type valves and fixed-speed motors. The applications of course must be matched, as would be the case with any design. A concomitant benefit is the reduction of one control element, i.e. a mechanical valve positioner for each loop. While we are on this subject, up front design of pumping systems is critical to long-term energy reduction due to the gross expense of replacing inefficient hydraulic systems.

Compressed Air

The typical specification for compressed air has a broad tolerance. This leads to inefficiencies arising from leaking systems and improperly high or low pressure settings. Poorly maintained line filters can introduce moisture or compressor lubricants and other undesirable elements into a process that might also adversely affect energy consumption. Air compressors also create waste heat that should be integrated whenever possible, but is often overlooked.

Hydraulic Systems

Piping is one of the most commonly overlooked system elements regarding energy. Typically, design engineers are forced to prioritize physical layout restrictions over all else. Still, overall pumping energy requirements can be modeled and used to at least identify potential energy savings.

Pneumatic Transport

Many chemical processing facilities transport solids by conveying them in air or nitrogen. Frictional losses are great in these systems and are supplied by compressors that themselves must be integrated into overall energy conservation plans.

Chillers

One of the largest consumers of plant energy is the chiller. Whether serving process heat dissipation loads or air conditioning work space, much heat can be recovered. Advances in technology have led to more efficient units that although costly, can lead to short payback periods, particularly at end-of-life replacement.

Electrical

- Motors—Choose energy efficient units and ensure they are matched to loads and set to turn off when not needed. Often running against closed valves can damage pumps and waste energy.
- Process Control—Modern Programmable Logic Controls (PLC) can allow a greater degree of flexibility than ever before. In fact, PLC reevaluation can lead to additional savings.
- Variable-Speed Drives—Do not use without proper design and selection, but can achieve significant savings.
- Variable-Speed Fans (VSF)—Air moving is another hidden energy gobbler. Often fans are either on or off, and use of VSF can save energy just as with VFD on motor driven pumps.
- Micro/Macro Electric Generating Turbines—Can be used to couple high to low quality heat source and sink.
- Power Topping Cycles—Particularly useful when multifuel capability or large combustible waste streams are available.

Water

- Once Through Cooling—Many facilities use water from rivers, wells, or municipal supplies. Often, the water is discarded after a single use—hence the term once through. Every attempt must be made to identify process or HVAC interchange opportunities. Lawn watering with this "Gray Water" is also a cost savings item if storage is possible. Avoid sprinkler operations during the rain by use of available rain detection equipment coupled to PLC.
- Process Feed Water—Many processes require water as feed and once through water can be substituted or augmented by clean water for this purpose. Interconnections are a bit more complex but should be considered in every conservation program.
- Cooling Towers—These can be used to reduce once through cooling water requirements. They must be maintained and can also be incorporated into Gray Water plans to save additional resources and money.

Electricity, Natural Gas, and Fuel Oil Supply

Periodic Cost Review

Review all energy procurement contracts to assure best possible pricing and structure. The US industry is undergoing a period of constantly changing de-regulation and restructuring, so these reviews should now as regular as those used for purchasing any other manufacturing supply.

Periodic Demand Charge Review

All electrical equipment briefly more energy during startup than while running at steady state. Utilities commonly levy an extra. "Demand Charge" for this. These charges can be avoided by load shifting to off hours, use of electrical load leveling technology, or just negotiations with the utility supplier. These charges can change even when base line rate does not and so require additional periodic review.

Waste to Energy

- Waste Energy and Material Evaluation.
- Review of all waste streams for energy recovery opportunity.
- Review of all waste streams for material recovery or waste exchange.

Facility Energy Integration

- HAVC Interconnection Review.
- Peak Demand Hours of Operation Review.
- Lighting.

PROCESS ENERGY CONSERVATION EXAMPLES

There is no end of examples, and we commend the reader to look at other sources, such as, the US Department of Energy. DOE maintains a website of process energy savings examples described later. We have provided several of these in the following article.

Chemical process engineering is comprised of physical and chemical manufacturing steps. The physical steps have become known as "Unit Operations" and are comprised of activities that make no change to the individual chemical constituents. Chemical manufacturing steps come under the category of "Reaction Engineering" that involve chemical changes to individual constituents. These are described in general below, and are followed by some common industrial examples.

Chemical Process Unit Operations

Distillation

This is the process of physically separating two or more components by the application of heat. The difference in component boiling point serves as the "separation factor" and as a result, this process uses both heat for boiling and cooling for condensing. As a result, this is one of the most energy intensive of all unit operations.

Liquid Extraction

Here pumping is often the largest source of energy and pressure drop through packed beds are the culprit. The main purpose of extractions is to transfer material from one stream to another, so mass transfer is the controlling design parameter. Continuous pumping of any type is a hidden energy drain so select low pressure-drop packing up front. Indeed this consideration is so new that certainly advances in packing may come about that will improve both energy and mass transfer.

Drying

This is an extremely common unit operation. In this application, controlled heat and humidity reduce or remove moisture from products. Wallboard and food products are two important examples. Low quality heat from other sources can assist in drying, while high moisture bearing waste hot air streams might preheat a process. Always remember surge requirements and design accordingly.

Freeze Drying and Crystallization

Ice undergoes a direct phase change from solid to gas. This process, termed sublimation, of course requires heat. Freeze-drying is employed where higher temperature evaporation is undesirable due to product breakdown or bacterial or enzymatic activity. Moisture is the primary culprit in food spoilage so this method greatly extends product shelf life. Refrigeration equipment is employed that should lend itself to heat recovery and integration, but must be coordinated closely with the process. In addition, vacuum pump equipment should also be included in energy optimization plans.

Crystallization

Water can be purified by alternating freeze-thaw cycles, taking advantage of the large difference in solubility of salt in water with temperature and phase change. As with drying, the focus will again be on refrigeration equipment and associated heat rejection.

Filtration

This process separates solids from liquids as well as liquids from other liquids. In solid-liquid filtration, pumping through a filter media causes retention of solids on the media. This is accomplished by either application of positive pump pressure or negative vacuum, both consuming energy. Liquids comprised of varying molecular size and structure components can be separated by ultrafiltration, a process that requires pump pressure and hence energy.

Chemical Process Reactor Operations

- Heat of Reaction, Exothermic–Heat is evolved and may need to be dissipated,
- Endothermic–Heat is absorbed and may need to be supplied.

Neutralization–A process of adding acid or base to produce a neutral product. These are generally exothermic, require heat dissipation and are ripe for heat recovery.

Polymerization–A very common process for convert individual monomer molecules into longer chain polymers with more desirable properties. Plastics are a good examples of polymers. Catalysts are often used and these processes are prime candidates for energy conservation.

Combustion–A very important class of chemical reactions, this is used to reduce non-recoverable organic waste materials to ash and heat through chemical reaction known as oxidation. Combustion is also employed to create steam and provide process and facility heat.

Whether exothermic or endothermic, energy integration is essential for modern reactor design. Most reactions have optimal temperature ranges so reliance on waste heat requires topping from an alternate source. Again, a modern programmable logic controller (PLC) can easily integrate two heat/cooling sources for on reactor.

Combined Heat and Power: Putting it All Together

Adsorption Chillers and Process Heat Recovery

These can be used to combine high and low quality energy to maximize efficiency. On their own, modern high efficiency chillers already save over older systems. Whenever the waste heat of compression can be recovered to provide process heat in, for instance, a distillation tower or dryer, the savings can often double that of the increased compressor efficiency alone.

Integrated Steam Generation from Waste Process Heat

Here again by combining low and high quality heat, maximum efficiency is obtained. In fact, not combining these two can lead to a decision not to invest as cost recovery time may be well beyond corporate policy.

Integrated Electric Power Generation from Waste Process Heat

Electric generation costs are comprised of nearly 33% line transmission losses. In addition, electric utilities do not always have a use for the low quality waste heat produced. Chemical process manufacturers should always explore the option at least of adding electrical topping cycles that are matched to process heat loads. Always remember to factor in demand charges as well when making investment decisions.

Whenever considering energy upgrades in Combined Heat and Power (CHP), remember that you can minimize break-even cost time for end of life equipment replacements, so make a special effort to integrate process and utility management groups fully in the capital equipment decision-making process.

Process Energy Examples

Distillation Energy Reduction

In this case, we examine 1000 Pounds/Hour of water distillation. This requires an approximately 1 million BTU/Hour heat source, typically a boiler, and a 1 million BTU per hour heat sink, typically a condenser. We will assume cooling water will be used in the condenser and that steam will be made in the boiler.

A common method of reducing distillation energy load is through mechanical vapor recompression (MVR). Simply stated, MVR places a compressor in the overhead vapors to raise the dew point significantly above the reboiler temperature. In this manner, heat is supplied entirely by the addition of work to the compressor through latent heat from overhead vapor condensation in the reboiler. In some instances, this can completely remove the need for an overhead condenser or cooler thereby saving either once through cooling water or reducing the load to the external cooling tower. Pre-heaters typically protect the turbine from harmful condensed liquid.

Another method, multiple effect evaporation (MEV), involves reducing the operating pressure in successive evaporator stages. With MEV, the same principle of MVR, i.e., condensing one stages' overhead against the boiling section of another stage is applied. In either case, the heat load is reduced by as much as 75%, with similar benefits to cooling water load. Of course, this introduces additional equipment cost and operational complexity and becomes part of the energy optimization equation. With a modern PLC, the control methodology to manage complex systems is now readily and inexpensively available—a fact that should make many retrofits of aging equipment very successful.

The above are common seawater desalination methods. They are also highly applicable to simple organic distillation purification systems. Typically, this is limited to systems where top and bottom tower temperatures are similar as with desalination and some solvent separations. As discussed earlier, MEB are highly coupled, and changes in energy flux within a distillation column will affect composition at both exits. The difficulty of process quality control will aid in screening distillation candidates.

Waste to Energy

Municipal Solid Waste (MSW) consists essentially of garbage, for lack of a better term. MSW in turn consists of environmentally recoverable materials (ERM) such as metal, paper, and plastic. MSW stripped of most ERM can still have a relatively high BTU content. Incinerating this material is not preferred environmentally as it requires extensive stack gas treatment and is not trusted by the public. Another process, pyrolysis can be used to produce a combustible gas, but a hybrid is generally preferred and is discussed below.

Classic pyrolysis involves heating organic material in the absence of oxygen. This process produces a combustible gas by-product, has very low particulate emissions, but produces a byproduct carbon char that still contains recoverable Btu. Newer methods incorporate a small, controlled quantity of air to minimize char production. The resulting medium-Btu gas produced can provide process heat through hot water or steam generation, or in some instances, a direct fuel supplement. The resulting pyrolysis ash is often disposed as MSW itself at greatly reduced volume, or in cement kilns where metal content is not a problem. The relatively low operating temperature allows the recovery of metals, including aluminum, as well as glass for recycling. Heat recovery from produced gas is also possible.

An additional benefit of this type of system may lie in its air permitting classification—i.e., not as an incinerator, as the extremely low flow of gases greatly reduces air particulate emissions common to incinerators. The low temperature also allows easier recovery of metal and glass waste. Pre-sorting is minimized, and materials come out clean and easier to recycle.

Fish Processing Filter and Evaporator

The processing of seafood produces many useful byproducts e.g., bone, protein, and oil. Much water is used throughout and often must be removed to recover these products. This is typically done with large evaporators and centrifuges that serve to concentrate the solids to a desired level. In addition, high-Btu-content oil is produced that can be burned to generate evaporator heat or electricity depending upon facility needs and layout. Large processors can produce thousands of gal/day of fish oil saleable for both nutritional value or blended for clean fuel power generation.

Waste Heat Recovery

Facility energy managers are in general familiar with methods used to intertwine HVAC components to recover waste energy from the various components on the utilities side of the house. The very same principles apply to interacting with chemical processes. In particular, be on the lookout for low quality heat, e.g., turbine extraction steam or reboiler condensate that is difficult to reuse in HVAC. Chemical processes often have stream preheating needs that can recover a high percentage of BTU content not thermodynamically available to work engines. HVAC, and air conditioning and plant operational energy integration is challenging, requires coordination and can be rewarding.

Checklist to Review for Process Energy Optimization

Avoiding startup and shutdown excursion is important in any process. During startup and shutdown, processes commonly run off spec and must be wastefully recycled. You may need to consider supplemental fuel to keep an electrical generator running when process output is low. Without a good relationship with utility providers, energy optimization can be difficult. This concept is generic to virtually any energy optimization scheme designed to capture non-steady waste energy. The key to success here is proper contracts with the utility and grid integration, onsite and offsite.

Review neighboring energy and raw material needs. You might be able to sell energy and chemical by-products just like any other commodity. It takes a little effort, but can be worthwhile in reducing permit and reporting requirements as well as attendant good neighbor status. A material is not a waste if it is still useful. Sometimes reclassification might be required, but EPA and its many State equivalents tend to be very supportive of all legitimate efforts to reduce waste production while conserving raw materials, and, as we have seen, energy.

Process Energy Checklist

- Establish Process/Utility Engineering Working Group—Necessary to ensure overall facility energy optimization.
- Conduct an MSW Audit—Make sure you are not throwing away Btu and recyclables such as metal, plastic, and glass. Get everyone on-board.
- Utility Bill Review—Identify and develop baseline costs for all process energy use including gas, electric, coal, etc.
- Determine Electric Company policy on purchase of onsite generated electricity.
  - Look for financial incentives for equipment purchase or Operations and Maintenance (O&M)
  - Utilize off-peak demand charge and rate structure to save on electrical operations
  - Consider micro-turbine power generation
- Identify Gas Company incentives
  - Look for financial incentives for equipment purchase or O&M
- Identify high quality thermal streams that might be directly tapped or easily upgraded to generate electricity.
- Utilize electric vehicles for plant operations and charge only at night when demand charges are lowest.
- Distillation and drying are high thermal process energy consumers that can often be modified to capture low quality heat, generate electricity, reduce cooling water demand or provide space heating. Multiple effect evaporation is an established method that should be explored.
- Waste Exchange Review—It is not waste until you declare it waste. If your process by-product is useable as a feedstock in some other company's process, then you have a product, not a waste. This will be saleable! If you have already declared by products a waste, they can always be "de-listed" with EPA. The process may

take some doing, but can reap financial and PR benefits as well.

- Load Shifting—Incorporate energy demand charge as function of time into all process design optimization and change programs. The City of Chicago constructed a refrigeration process that makes ice at night and provides district cooling during the day in downtown Chicago. Off-peak electric rates are usually quite a bargain. Factor this into any processing scheme.
- Leading Energy Indicators—Select metric feedback elements to monitor energy performance.

Office HVAC and Lighting

Although this topic is covered in greater detail elsewhere in this encyclopedia, it is worth discussing here as older incandescent bulbs produce more heat than light. Newer compact fluorescent bulbs are at least four times as efficient and last five times longer, saving on energy and maintenance with payback periods of a year or so. Replacement of old magnetic ballast T12 fluorescent bulbs with new style T8 or T5 with electronic ballasts can have payback periods of 2-3 years in upgrade mode or less than a year in the case of end of life replacement. Cooling and heating in large buildings is often accomplished with boilers and chillers with mixing of air taking place in air handlers. Make sure you are not balancing steam against chilled water to maintain office or plant air temperature set points. Several basic ideas to use in saving energy in large or small facilities include:

- Use area motion detectors to actuate light and in some cases zone HVAC if available.
- Use setback thermostats programmed for heating and cooling for several seasons.
- Consider off site control via telephone or internet of HVAC and lighting for large buildings.
- Store process gray water for lawn sprinkling.
- Examine process heaters and chillers for interconnection, at least for topping service.

Building Energy Management and Improvement

Often overlooked in industrial environment, this is area is ripe for energy integration with low quality waste process heat or excess refrigeration for cooling. The off peak ice machine concept can also work here to help reduce demand charge. Consider replacing older copiers and laser printers that require high temperature for operation for replacement with energy star labeled products. The EPA has computer programs to quantify equipment replacement benefits. Computer energy conservation modes must be set by corporate IT departments to assure the proper sleep or hibernation required to save energy when machines are left unused for even a relatively short time. Screen savers of the past do not save energy. The speed of modern computer processors and hard drives makes hibernation or sleep modes very effective. Newer chips can also be set to run slower, consume less energy, and produce less heat. Photocopiers are huge energy wasters and must go to sleep within a reasonable time. Be careful not to shut down important elements such as fax and network printers now common in multifunction office machines.

Government Energy-Related Programs

DOE Programs

The DOE has industrial technology program (ITP) support available. Elements of this program include free technical assistance and project co-funding opportunities. The program identifies four broad areas in the Save Energy Now arena and can provide free energy audits to facilities around the United States. These audits assist companies in charting facilities energy use, identify areas of concern by matching energy bills with similar facilities and recommending changes along with estimated pay back periods to allow project prioritization.

EPA Energy Star Program

EPA and DOE have interlocking missions here, but EPA manages this program for both agencies. The program includes product labeling with energy use and efficiency, electronic automatic energy reduction for computers and other related class of electronics as well as industrial integration. Websites of both offer consumers and energy engineers alike access to a wealth of useful and growing information.

In our article on “Energy Savings and Optimization Case Studies,” which follows this one, we have selected several representative cooperative industrial examples from the US Department of Energy Web Site.

Six Sigma Methods: Measurement and Verification

Carla Fair-Wright
Maintenance Technology Services, Cameron's Compression Systems Division, Houston, Texas, U.S.A.

Michael R. Dorrington
Cameron's Compression Systems Division, Houston, Texas, U.S.A.

Abstract

"If you can't measure, you can't control—if you can't control, you can't manage—if you can't manage, you can't contain costs." As a process-oriented activity, energy management begins with a thorough assessment of use and culminates in the implementation of an effective energy management plan. Applying the five process improvement steps of a Six Sigma DMAIC (Define, Measure, Analyze, Improve, and Control) activity provides a way to define, measure, analyze, improve, control, and verify energy use and savings. The straightforward approach of this process provides a cost-effective means of optimizing facility energy and cost efficiencies.

These efficiencies are transformed into energy savings based on improvements to both facility process and utility-related activities, with process-related activities encompassing all product manufacturing events, and utility-related activities including automated building systems and all other activities not directly related to manufacturing events.

The DMAIC process for each begins with a well-defined problem statement that is supported by examination of existing data to formulate the project premise. Combining "tribal knowledge" with sound data analysis highlights the pathway to process improvement and long-term savings.

NOMENCLATURE

DMAIC Acronym for Define, Measure, Analyze, Improve, and Control, a five-phase, Six Sigma project-focused approach to continuous process improvement

Tribal knowledge Tacit knowledge held by individuals or groups based upon their individual and collective experiences

Stage gates A management review of the process used to screen and pass projects as they progress through their various stages of development

ROI Return on investment = net income/investment, where net income = expected earnings

MSCF Million standard cubic feet

Energy audit A systematic survey conducted to measure and record energy consumption, used to identify opportunities to reduce energy use and cost

Process capability ratios Capability indices used to measure short-term and long-term process standard deviation

Special-cause variation The rate or magnitude of change in an output resulting from an unexpected event

Common-cause variation The rate or magnitude of change in an output resulting from expected or everyday events

Run chart A graphical display of data representing process outputs over a given period of time, used to determine stability of a process

Process variation The quantifiable difference between individual measurements, represented in a designed sequence of operations or events

DPMO Acronym for defects per million opportunities

Project charter A formal document that defines the mission of the project, project scope, financial benefit, measurement metrics, time frame and resource requirements

As-is process map The current flow of events in a process containing the inputs and outputs of each process step

Capability analysis A measure of how a process is performing, visually represented in terms of a bell-shaped curve, in terms of how much variation exists in the process and how centered it is within the limits of the process

Keywords: DMAIC; Tribal knowledge; Stage gates; ROI analysis; MSCF; Energy audit; Process capability ratios; Special-cause variation; Common-cause variation; Run chart; Process variation; DPMO; Project charter; As-is process flow; Capability analysis; Cause-and-effect matrix; Failure modes and effect analysis; Risk priority number; Measurement system analysis; Graphical analysis; Hypothesis testing; Regression analysis; Cost-benefit analysis; Predictor variable; Response variable; Correlation coefficient; Coefficient of determination; *p*-value; KPIV; KPOV; Control plan.

Cause-and-effect matrix A chart used to rank the relationship between process inputs and outputs of a designed sequence of operations or events

Failure modes and effects analysis FMEA; a document that provides a systematic technique to identify and analyze potential failure modes by quantifying them according to the severity of risk they present

Risk priority number RPN; a numerical indicator used to identify the severity of risk

Measurement system analysis MSA; an evaluation used to determine the precision of a given response

Graphical analysis An investigation of how a process or event is performing by representing the data in diagram form

Hypothesis test A statistical method to determine if there are differences between two or more process outputs

Regression analysis A statistical technique used to investigate and model the relationship between dependant and independent variables

Cost-benefit analysis A technique used to analyze competing alternatives based upon comparison of total cost to total return

Predictor variable A value that can be used to predict the value of another variable

Response variable A value that represents the outcome of a specified process or experiment

Correlation coefficient r; statistic that describes the strength of a relationship between two variables

Coefficient of determination r^2; the square of the correlation coefficient, and a measure of the strength of the association between two variables

***p*-value** Probability value; a number that reflects the probability that a statistical result happened by chance, causing a hypothesis to be either accepted or rejected

KPIV Acronym for Key Process Input Variable

KPOV Acronym for Key Process Output Variable

Control plan A document that ensures processes are operated and monitored consistently so that the desired result meets the defined requirements

Cp Process capability; represents the spread of the process. A Cp Value of 2 is representative of a 6σ process (0.002% rejects)

Cpk Process capability of non-centered process; is an indication of how the process is aligned to the process mean. Cp and Cpk are equal in a centered process

INTRODUCTION

We live and work in an ever-increasingly volatile and complex business environment that has begun to recognize energy use and efficiency as strategic elements of our overall business planning process. Under deregulation, energy providers pass on a higher percentage of energy cost increases to the consumer in the form of cost adjustment factors. These added costs, the efficiencies of the energy conversion processes we employ, and environmental effects are what Six Sigma practitioners would define as critical inputs (*X*-values). Outcomes—or *Y*-values—are a function of these critical inputs and stated in terms of our monthly utility expense. By applying Six Sigma methodology to the issue, we are able to define energy use as a process and employ a useful set of metrics to measure, improve, and control variation in our monthly utility bills.

ENERGY MANAGEMENT—PURPOSE AND GOALS

The goal of energy management planning is to optimize both the consumption energy and its cost. This article offers an overview on how to apply Six Sigma methodology to continuously improve energy consumption processes using DMAIC principles. Six Sigma methodology also benefits energy managers by adding stage-gate approvals into the energy management plan development process.

SIX SIGMA—BACKGROUND

Established in the 1980s at Motorola and brought into mainstream use by GE in the mid-1990s, Six Sigma employs data-based decisions built around baseline measurements to reduce waste and process variation. This fundamental objective is often achieved using the five-step DMAIC process (Define, Measure, Analyze, Improve, and Control). Since most energy use projects are site specific, this article will serve to guide users in the application of Six Sigma methodology rather than developing a specific case study. Additionally, sample sets of data have been employed to simplify data analysis while offering an acceptable level of confidence in the results.

FINANCIAL CONSIDERATIONS AND INCENTIVES

Capital expenditures for facility energy infrastructure projects are rated based upon financial criteria that seek to establish their expected rate of return. While financial analysis of large energy infrastructure projects is often complex, smaller straightforward projects generally stress the expected return on investment (ROI). This article

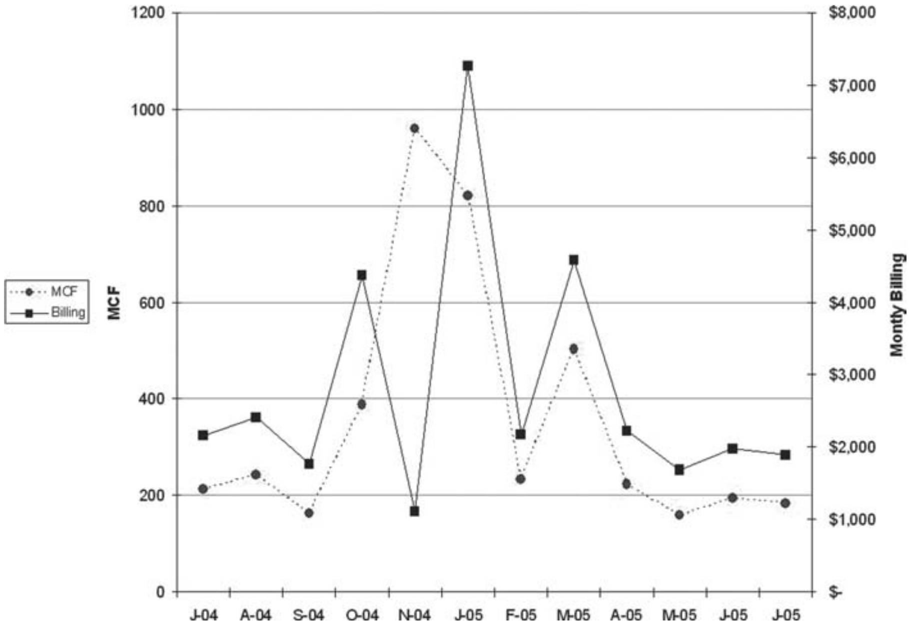


Fig. 1 Energy usage.

employs ROI Analysis to forecast the expected rate of return.

It should be noted that the Energy Policy Act of 2005 provides substantial tax incentives to encourage improving energy efficiency in both new and existing buildings. A major aspect of the act is the ability of facility managers to obtain new energy equipment and apply any derived cost savings to energy utilization.

DATA SOURCE

Research data used in this article were obtained from the monthly service billings of a small aftermarket repair facility in Houston, TX., that employs a single analog utility meter to record energy consumption (Fig. 1). This facility was selected based upon the simplicity of its metering scheme and the ability to readily obtain a sample set of data that fit the reporting purposes of this article. Larger facilities will employ sub-metering or data-recording devices that allow users to categorize energy use by process segment. The ability to link energy use to individual machine tools or processes enhances the ability to analyze data and verify the impact of specific improvements.

In practice, one of the first steps taken in evaluating energy use is an energy audit. Usage data are gathered from a variety of sources including metering devices, sensors, and utility bills, and combined with maintenance information to identify energy cost reduction opportunities. Self-assessment and assessments conducted by university-based industrial assessment centers are two low-cost means of performing an energy audit.

SETTING ENERGY CONSUMPTION TARGETS

Determining the need for improvement requires establishing current usage (baseline) data that can then be compared with desired objectives or best industry practices. Performance against the desired outcome can then be measured in terms of its process capability ratios (Cp and Cpk).

SIX SIGMA—THE DMAIC PERSPECTIVE

Overview

As a rigorous methodology that uses data-based decisions and statistical process control, Six Sigma provides firms with a robust platform to minimize variability, defects, and waste in the processes we employ to convert goods and services into revenues. In practice, Six Sigma represents how well a process is performing and how often a defect is likely to occur—the higher the sigma value, the better the process is performing.

Since energy savings programs are essentially project activities, we can readily apply DMAIC methodology to characterize and optimize desired benefits. The five phases of this process are as follows:

- Define—Stating the desired need in the form of a problem statement.
- Measure—Determining the baseline and target performance values.
- Analyze—Relating process inputs (critical *X*'s) to process outputs (*Y*'s).

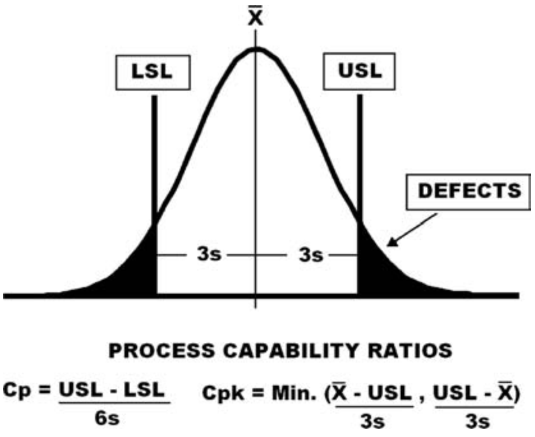


Fig. 2 Process capability.

- Improve—Identifying improvements to optimize outputs.
- Control—Establishing control procedures and accountability.

Variation and Mean Shift

Every process includes both common-cause variation and special-cause variation. Common-cause variation occurs naturally in any process, while special-cause variation is nonrandom and the result of an action or a series of actions. Process improvement seeks to reduce special-cause variation, shifting the mean to center and stabilizing the process (Fig. 2). The capability of the process to consistently perform within specification limits relies on achieving both. A common tool employed to display both stability and variation is the run chart, which allows users to quickly gauge the amount of variation in a process. Points outside the process upper and lower limits, unequal dispersions around the mean, and pattern variations are all indicators of a process variation requiring further investigation.

Defect Prevention

A process, product, or result may contain a single or multiple number of defects before it is considered defective or out of specification. Defect prevention results from error proofing the activity. The common defect measure is expressed in terms of parts per million or defects per million opportunities (Table 1). A process operating at Six Sigma levels produces no more than 3.4 defects per million operations, meaning that the process is operating at a 99.99966% defect free rate. For energy savings projects, a defect may represent any point outside of the run chart control limits.

Data-Based Decisions

Six Sigma converts data into useful information that allows users to base their decisions on facts rather than opinion. Data-based solutions may result in improvements that match expert opinion, but data-driven solutions coupled with statistical process control remove errors while adding process control capability—all resulting in higher probability of a successful outcome.

DMAIC METHODOLOGY

Define

During this initial project phase, the problem statement is created in the form of a project charter. Project charters establish the project's scope and improvement objective; detail the financial impact of the project in a business case; and forecast the project's expense, savings, and required resources.

Measure

Understanding the Process

Before a process can be improved, it must first be measured to establish a baseline for comparison and targeted improvement objectives. Baseline data (typically in kilowatt-hours or million standard cubic feet) may come from utility bills, meter readings, or data recording devices. Following data collection, key process variables are identified by creating a top level as-is process flow to help narrow the focus of the improvement effort to those key process steps having the highest levels of special cause variation.

Process Capability

The demonstration of process performance is accomplished through capability analysis, by comparing

Table 1 Defects per million opportunities yield and Sigma

Sigma (σ) value	Accuracy (%)	Defects per million opportunities (DPMO)
1	30.85	691,500
2	69.15	308,500
3	93.32	66,800
4	99.38	6,210
5	99.977	233
6	99.999	3.4

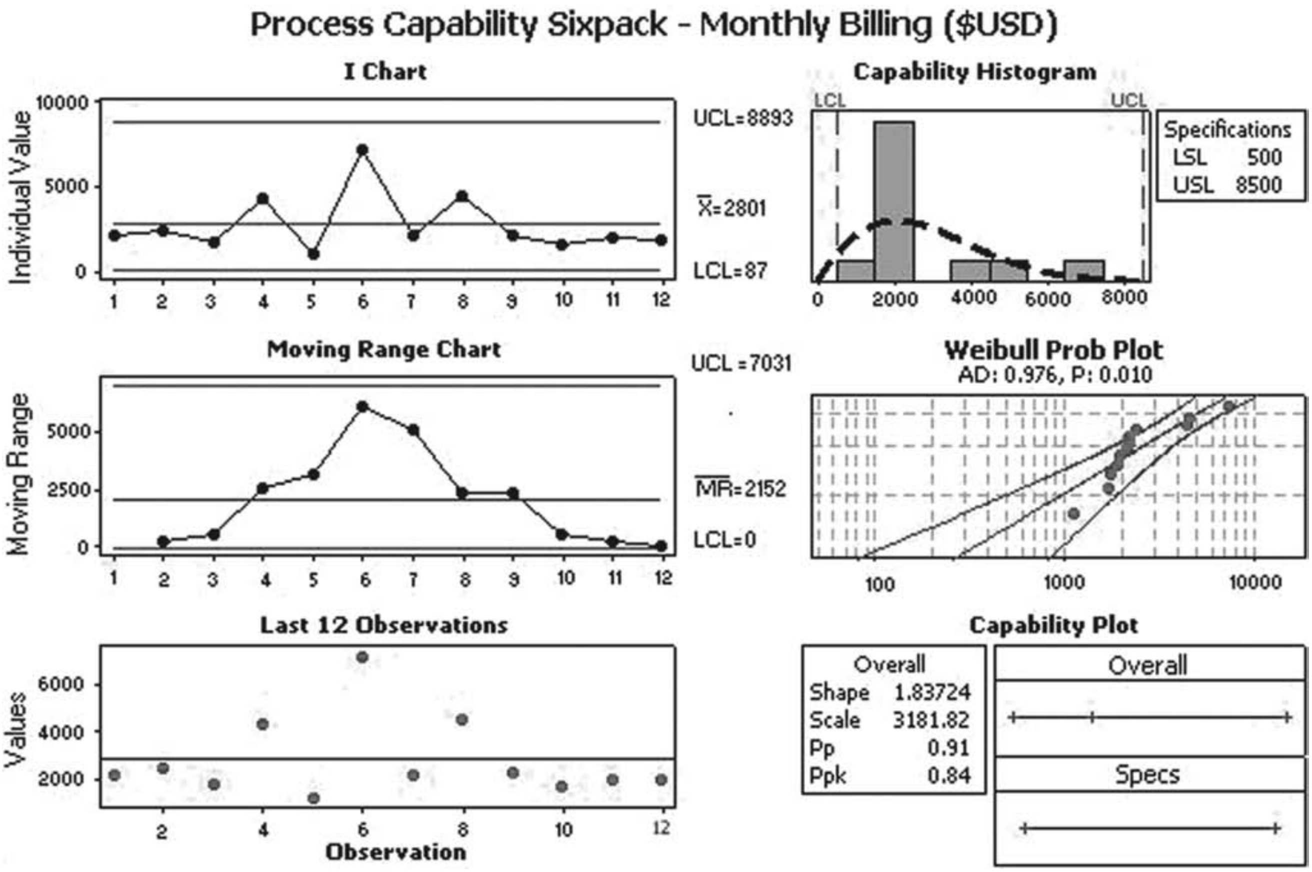


Fig. 3 Energy usage process capability.

the width of the process variation to the width of the acceptable process tolerance limits, or by measuring the defects present in the existing process. The standard measures of capability are defects per million opportunities (DPMO) and process capability ratios (Cp and Cpk).

Analyze

Non-Normal Data

Before performing any data analysis, one should check data for normality, since standard capability indices contain an assumption of normality. With energy usage programs, data is likely to be non-normal due to a number of factors including monthly utility fees and surcharges. The data used in this article is non-normal data that was transformed before use (Fig. 3).

One strategy is to make non-normal data resemble normal data by using a Box-Cox transformation function to convert the data closer to a normal distribution (Fig. 4). The underlying technique of the Box-Cox transformation is based on developing a normality plot using the correlation coefficients for the defined transformation parameter (λ, the lambda value). The relatively high *p*-value taken from the fitted line plot on Fig. 4 indicates that the transformed data is normal and capable of being analyzed using Six Sigma statistical tools. The capability charts indicate that significant variation occurred between periods 3 and 10, signifying that further investigation of these eight time periods is required. These periods should be analyzed with reference to the process step inputs and outputs developed in the as-is process map to identify any observable key input or output variables.

Sources of Variation

Once as-is process capability is defined, it is necessary to understand cause and effect relationships. A cause-and-effect matrix is the tool used to rank relationships between the key input and output variables. When complete, the cause-and-effect matrix is linked to a failure modes and effect analysis (FMEA) by transferring the highest-ranked input variable or variables onto the FMEA. This analysis is used to identify and force rank by criticality, the potential failure modes, and their effects and causes, using a risk

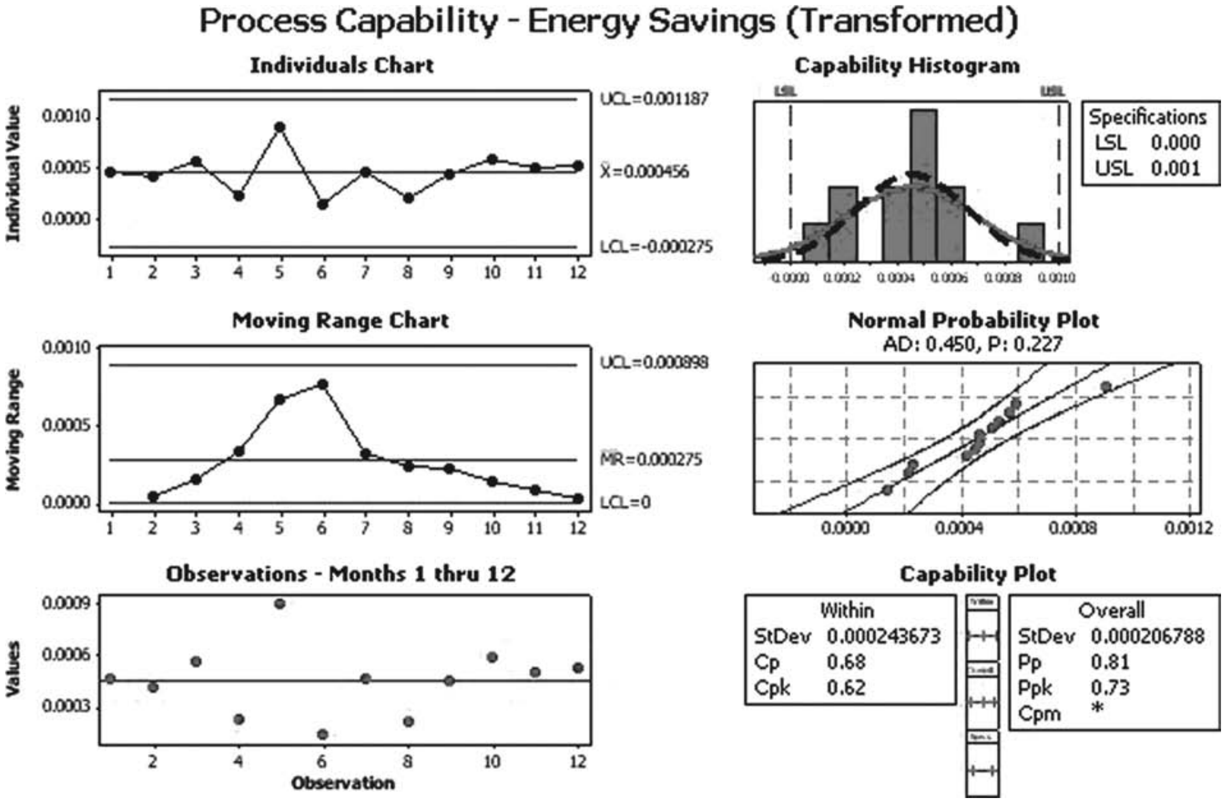


Fig. 4 Energy usage process capability transformed.

priority number (RPN). Only the highest-rated factors qualify for investigation.

Measurement Error

Accuracy of the measurement system can be the predominant reason for process variation. Employing measurement system analysis (MSA) will allow the firm to evaluate the accuracy and precision of the measurement system. Where measurement systems are inadequate, it is recommended that they be upgraded to a level consistent with the project requirements.

Corrective Actions

Completion of the FMEA requires statement of the corrective action needed for process improvement. For each action taken, it is necessary to recalculate the RPN to validate process improvement. A variety of Six Sigma statistical tools are available to confirm the impact of the corrective actions, including graphical analysis, hypothesis testing, and regression analysis. As a supplement to the FMEA, a cost-benefit analysis should be developed to define the optimal level of planned investment.

IMPROVE

Improvement seeks to optimize outputs (Y's), eliminate or minimize defects and variation, and statistically validate new process operating conditions.

Variable Relationships

The governing principle behind variable relationships is expressed in terms of $y=f(x)$, wherein key output (y) is stated as a function (f) of the key inputs (x). A common statistical tool used to examine this relationship is regression analysis, which correlates the degree of association between the predictor variable (x) and the response variable (y) in terms of r, the correlation coefficient. Where $r>0.8$ or <-0.8, a significant relationship exists; for values between -0.8 and 0.8, no significant relationship exists. The amount of variation in the output that is explained by the regression model is R^2, the coefficient of determination. In terms of energy analysis, a strong correlation indicates that the output (y) or energy use is defined by the input values (x) being modeled. The final term dealing with correlation in the regression is the p-value, which indicates whether there is a correlation between the variables. A p-value can also be used to determine normality, which can be seen by the fit of the variables to a fitted line plot (Fig. 3).

For the facility cited in this article, it was determined that utility fees and surcharges were the critical inputs. Regression analysis indicated a weak correlation to the various input factors of shop load, utilization factors, maintenance, and seasonality, due in part to the limited amount of data and unpredictable increases in utility energy fees and surcharges. Based upon the age of the facility and a cost-benefit analysis of the upgrades needed to operate the facility at optimal energy levels, it was determined that the best alternative was to switch to a lower-cost energy provider as the means to reduce energy costs.

New Process Capability

Following implementation of the recommended actions, it is necessary to establish the new relationships of key process input variables (KPIVs) and key process output variables (KPOVs) by running a second capability analysis using data from the improved process. This information is then compared to the original baseline data to establish the level of improvement obtained. The final steps in defining the new process capability require updating the process map to reflect implemented changes and validating project cost savings.

CONTROL

The final project phase documents, sets up control criteria and assigns accountability for sustaining the gains achieved through process improvement. Control plans are intended to permit operating processes consistently on target with minimum variation by defining and monitoring critical success factors (KPIVs and KPOVs); as well as the employed measurement methods, frequency, and decision rules governing corrective action. It is important to remember that the control plan is a living document and part of a continuous process improvement activity.

CONCLUSION

The primary advantage of Six Sigma as a continuous process improvement tool, is its ability to provide a set of metrics that can then be used to understand and correct process variation. Six Sigma is a project-focused activity that adapts well to time-dependent activities. The five-phase DMAIC model is commonly applied to characterize and optimize industrial processes. Understanding how energy is consumed in the processes we employ requires that we first establish a method of measurement to identify how energy is being consumed and in what quantities. Organizing data into useful information is accomplished in the measurement phase of the DMAIC process, which establishes the baseline against which all future improvements will be compared. What prevents decline of the improved process is the control plan, which is established to identify ongoing surveillance requirements and corrective action plans.

RESOURCES

The American Society for Quality http://www.asq.org/

The DOE Building Technologies Program http://www.eere.energy.gov

International Performance Measurement and Verification Protocol http://www.ipmvp.org

BIBLIOGRAPHY

1. Watson, J.J. Satisfaction through six sigma. *Engineered Systems*, March 2003; 1–2. http://www.findarticles.com/p/articles/mi_m0BPR/is_3_20/ai_98780226 (accessed November 2005).
2. Muller, M.R.; Papadaratsakis, K. *Self-Assessment Workbook for Small Manufacturers*. Version 2.0, Department of Energy (October 2003).
3. The Black Belt Memory Jogger™ *Six Sigma Academy*, 1st Ed.; 2002.
4. Introducing Information Systems for Energy Management, GPG 231, Action Energy, http://www.thecarbontrust.co.uk/energy/pages/home.asp.

Solar Heating and Air Conditioning: Case Study[☆]

Eric Oliver
K. Baker
M. Babcock
EMO Energy Solutions, LLC, Vienna, Virginia, U.S.A.
John Archibald
American Solar, Inc., Annandale, Virginia, U.S.A.

Res–Solar

Abstract
Mechanical cooling is one of the largest end uses of electricity, and it contributes significantly to peak demand and on-peak energy consumption. The ability to offset electricity consumption during peak usage periods by eliminating mechanical cooling can save costs, reduce strains on transmission grids, and free up much-needed peak period capacity for electric generation utilities. Solar energy use, long encouraged for offsetting space heating and daylighting energy consumption, can now be utilized for offsetting cooling consumption. By using solar heat to regenerate a desiccant dehumidification system, and utilizing direct and indirect evaporative cooling to condition superheated dry supply air, occupied spaces can be cooled using no grid energy input. A demonstration project has been designed to cool an inspection station at the U.S. Pentagon in Arlington, Virginia, and is expected to provide 65°F, moderate humidity air at summer design conditions.

INTRODUCTION

An inspection station at the Pentagon Central Plant currently uses a through-wall heat pump for cooling and heating. This station was chosen for a demonstration project to test the application of space cooling using only solar thermal, desiccant dehumidification, and indirect evaporative cooling. By using collected rainwater for evaporative cooling and photovoltaic panels with battery back-up to run peripheral equipment, this system will have the potential to operate completely off grid, with no purchased electricity or associated emissions. Cool air will be provided directly to the station using a four-stage process:

- Stage 1 : An American Solar Inc. tile roofing system will be used to heat ambient air up to 200°F.
- Stage 2 : Supply air for the station will be dehumidified and superheated using a small desiccant wheel. The wheel will be regenerated using the solar heated air.
- Stage 3 : The superheated air will be cooled using a direct air–air heat exchanger with ambient air.
- Stage 4 : The hot air will be further cooled using two stages of indirect evaporative cooling, first with saturated ambient air and second with a portion of the supply air diverted into a second indirect evaporative cooler.

Based on a psychrometric analysis, it can be shown that 65°F 60% relative humidity (RH) air can be delivered from Washington, DC, design ambient conditions of 85°F dry bulb (DB)/75°F wet bulb (WB).

COOLING SYSTEM METHODOLOGY

The cooling process designed utilizes three separate air streams: the supply or primary air stream, the secondary air stream, and the regenerative air stream (RAS). Further details of each of the equipment used will follow later in this abstract. Fig. 1 is a schematic of the design.

The primary air stream is drawn from outside air through air filters and is mixed with the exhaust air from the second evaporative cooler (IEVAP-2). This mixing process reduces the enthalpy and temperature of the incoming air during design conditions. The primary air then passes through the dehumidification side of the desiccant wheel (DESC-1). The air is then passed through an air-to-air heat exchanger (HX-1) with the regenerative air to cool the primary air. In the next stage, the primary air is further cooled by passing through an indirect evaporative cooler (IEVAP-1). Finally, the primary air passes through a Maisotsenko-type indirect evaporative cooler (IEVAP-2) to provide the final conditioning for the supply air. During this final stage, the primary air is divided into the primary air stream and an exhausted secondary air stream.

[☆] This entry originally appeared as "Solar Heating and Air Conditioning at a Pentagon Security Inspection Station" in *Energy Engineering*, Vol. 101, No. 4, 2004. Reprinted with permission from AEE/Fairmont Press.

Keywords: Heating, Ventilating and Air Conditioning (HVAC); Solar heating; Cooling; Case studies.

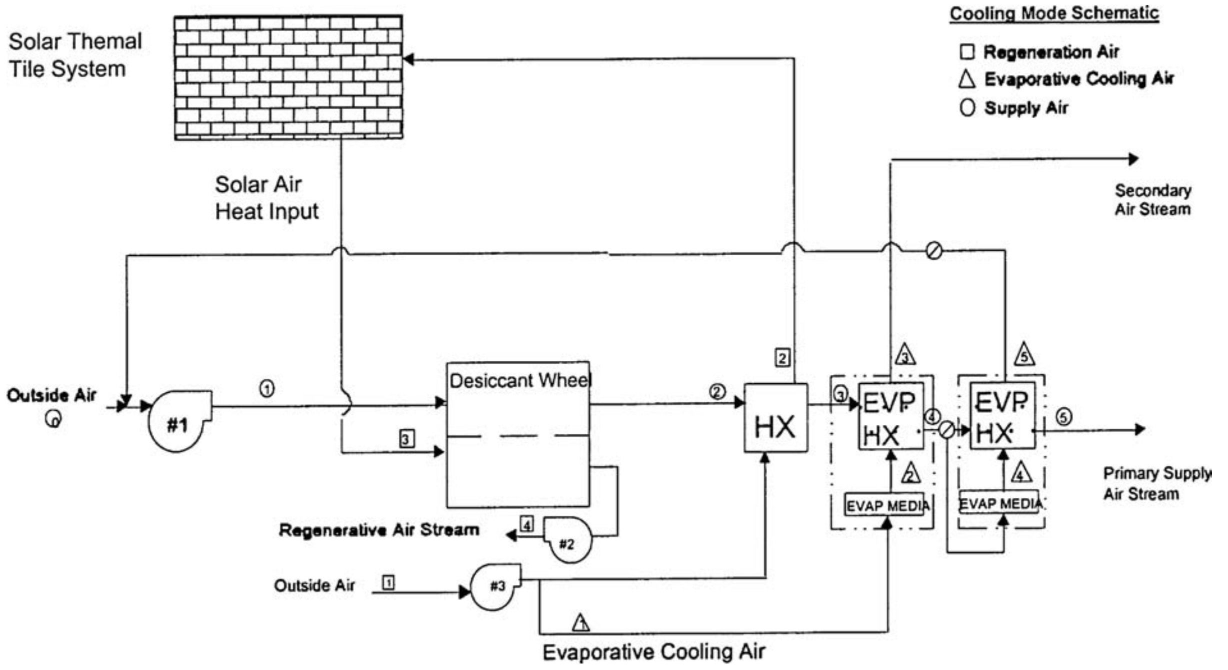


Fig. 1 System schematic.

The secondary air stream feeds IEVAP-1. The secondary air stream is intended to maximize the removal of sensible heat from the primary air and thereby reduce the primary air temperature. To this end, the airstream is designed with twice the airflow as the primary air.

Similar to the primary air stream, the RAS is drawn through air filters from outside. This airstream is first passed through HX-1 to recover heat from the superheated and dried primary air stream. This preheated air is then sent to the solar roof tile system to capture the solar energy and raise the RAS temperature. The solar heated regeneration air passes through the regeneration side of the DESC-1 to enable the desiccant to release and exhaust the absorbed moisture into the air stream.

PSYCHROMETRIC ANALYSIS

Fig. 2 illustrates the psychrometric path of the primary air stream. It begins at State 1—a high temperature, high humidity, and high enthalpy condition. The exhaust from the IEVAP-2 is saturated with water vapor, but has a lower temperature-humidity ratio and enthalpy than the outside air. The mixing of the airstreams reduces the overall enthalpy at State 2. State 3 occurs after the desiccant wheel; as expected, the moisture content is greatly reduced while the temperature is significantly increased. It is important to note that the total enthalpy of the air is actually increased from State 2 to State 3. States 4–6 are the result of sensible-only cooling through the heat exchange with other air streams in HX-1, IEVAP-1, and IEVAP-2. The overall result is a strategy that first shifts the enthalpy of the airstream from latent to sensible energy, and then uses heat exchange with original or tempered outside air to remove the sensible energy and produce cool air without the mechanical cooling.

As Fig. 3 shows, the psychrometric path for the RAS is simple. It starts at State 1, the outside air conditions, and gains sensible heat from the heat exchange at HX-1 to arrive at State 2. The air stream is further heated sensibly by the solar irradiance captured from the solar tile system, and reaches the regeneration temperature at State 3. Once the RAS has passed through the desiccant wheel, State 4 represents the cooler, wetter exhaust air containing the additional moisture from the desiccant.

SOLAR ROOF TILE SYSTEM

American Solar, Inc. has patented a technology to provide solar-heated air using a tile roof system. The solar thermal tile system is designed to provide dual functionality: to serve as the roof of a space and to provide solar heated air. This system translates a portion (~35%) of the solar insulation available in a given region into preheated air, which can be utilized for heat exchange, thermal storage, space heating, and in this case, desiccant regeneration air. Radiant energy is absorbed by black-painted metal absorbers and heats the air surrounding the absorbers. Transparent tiles over the absorbers contain the heated air, and allow a fan system, at a plenum at one end of the roof, to collect the solar-heated air.

Stagnation tests show that a typical system can achieve internal air temperatures of >200°F (94°C) and more than 130°F (72°C) above ambient temperature during the

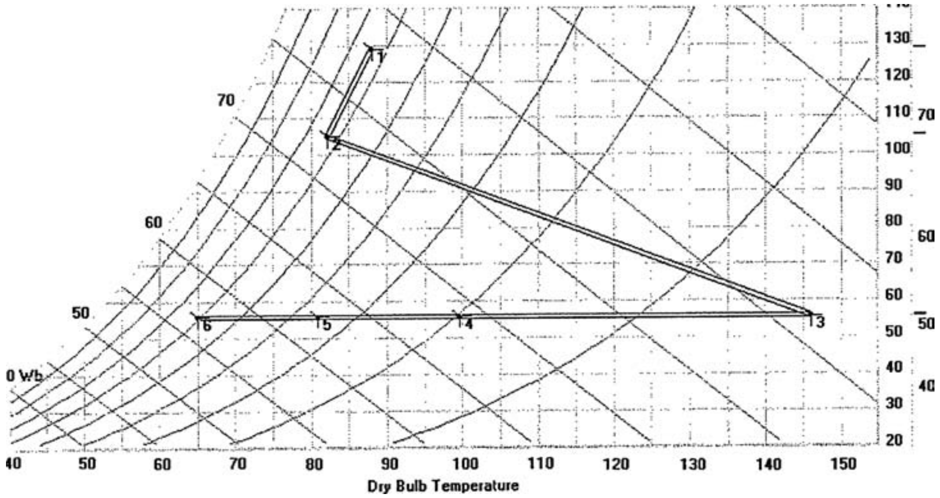


Fig. 2 Supply air stream psychrometric path.

summer months. An air flow test with an early prototype showed outlet air temperatures of 160–180°F (71–82°C) are possible. Higher temperatures are expected with optimal orientation, improved materials (i.e., absorber, glass tile, etc.), and optimal air flow given a system's configuration.

For this particular application, the preheated solar air generated from the solar roof will be used to regenerate a high temperature substrate desiccant, which will be utilized to dry the primary air stream. To achieve the temperatures need to regenerate this particular desiccant, the complete solar roof system had to be designed in three stages:

- Stage 1 : This roof section includes about 150 ft^2 of dark-colored standing seam panels, set at a tilt of 22.9° and facing north. The panels will heat up, and preheated air from underneath the seam panels will flow to the next roof section.
- Stage 2 : This roof section includes about 150 ft^2 of the ASI tile system, set at a tilt of 22.9° and facing south, that will boost up the temperature coming off the standing seam panels and deliver it to the next roof section.
- Stage 3 : This roof section includes about 90 ft^2 and faces south with an angled reflecting surface opposite it (to provide additional insulation). Stage 3 takes the incoming air from Stage 2 and increases the temperature of the air further. The air from Stage 3 is then delivered to the desiccant wheel for regeneration.

DESIGN AND CONSTRUCTION STRATEGY

The overall goal of this showcase project was to create an innovative and fully integrated cooling system which would also have the capability to provide solar-heated air for space heating in the winter. To provide power for peripheral equipment, such as fans and the water pump, several photovoltaic shingles will be installed in the roof system. This particular application will include battery a

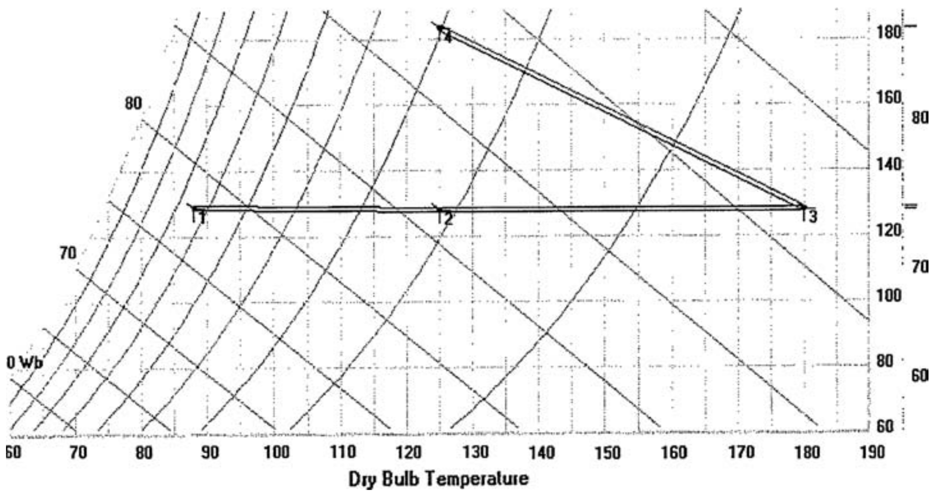


Fig. 3 Regeneration air stream psychrometric path.

backup and backup grid connection for battery charging, but the showcase is meant to demonstrate the potential for off-grid applications. Highlighted design features and modifications for test bed purposes include:

- The ability to alter and test the impact of different face velocities across the solar roof to increase the regeneration air temperature.
- The ability to adjust the relief damper position for the regeneration air to properly pressurize the solar roof.
- The end user's ability to adjust the temperature when the solar cooling system is not providing adequate supply air temperatures during muggy days or when there is little to no (night) solar isolation available.

Some of the challenges and system constraints include:

- The minimal amount of space available for routing ducts and mounting heating, ventilating and air conditioning (HVAC) equipment underneath the solar roof and above the 9′ carport ceiling limit.
- The integration of the solar system with the evaporative cooling system.
- The integration of the PV shingles and the ASI tile system into a single entity.
- Minimizing pressure losses through the supply system to minimize fan power consumption and energy losses.

CONCLUSION

There is a great potential to provide spot cooling for remote locations using solar-regenerative desiccant dehumidification, heat exchange, and indirect evaporative cooling. The economics of such an installation will be greatly dependent on climate, energy rates, benefits of avoided grid extension, and other factors where adding mechanical cooling to a location is costly. The potential impact on operating costs, utility generation and transmission demand constraints, and environmental improvement from reduced emissions is significant.

BIBLIOGRAPHY

1. ASHRAE 1993 Fundamentals.
2. ASHRAE 2000 HVAC Systems and Equipment.
3. Watt, J.R.; Brown, W.K. *Evaporative Cooling Handbook*, 3rd Ed.; Fairmont Press: Lilburn, Georgia, 1997.
4. Archibald, J. *A New Desiccant Evaporative Cooling Cycle for Solar Air Conditioning and Hot Water Heating*, Proceedings of the ASES Annual Conference, 2001.
5. U.S. Patent #5,651,226 to Archibald dated July 29, 1997.
6. Coolerado Coolers, www.coolerado.com
7. Novelaire Technologies, www.novelaire.com

Solar Thermal Technologies

M. Asif
School of the Built and Natural Environment, Glasgow Caledonian University, Glasgow, Scotland, U.K.

T. Muneer
School of Engineering, Napier University, Edinburgh, Scotland, U.K.

Abstract
Solar thermal technologies are amongst the most diverse and effective renewable energy technologies. They range from low-temperature ($<70°C$) and simple in operation technologies such as solar space heating and solar cooking to high-temperature ($>200°C$) and sophisticated ones like solar air conditioning and solar thermal power generation. Solar thermal technologies, similarly, have a broad economic bandwidth amongst them. The most successful application of solar thermal technologies so far has been in the form of solar water heating. In this entry, a brief overview of the prominent solar thermal technologies has been presented in which the technology fundamental and operational principles of these technologies have been precisely discussed.

INTRODUCTION

The sun is a source of energy on Earth, and in a sense, it is a source of life; as energy is the most important commodity for life besides providing light and heat to the planet. The reaction between the sun's energy and Earth's atmosphere determines weather patterns and rainfall, and Earth's tilt towards the sun creates the seasons. Its role in photosynthesis helps plants to grow and its role in biodegradation helps complete the natural cycle of ecosystems. The sun also sources several other forms of energy on the planet: wind power depends on the sun's impact on atmospheric movement as it creates wind patterns; through photosynthesis, sun contribute to bioenergy (wood and other organic materials); and fossil fuels indirectly owe their creation millions of years ago to solar energy.[1]

Solar energy is one of the most promising renewable technologies. It is abundant, inexhaustible, environmentally friendly, and widely available. Solar energy has the potential not only to play a very important role in providing most of the heating, cooling, and electricity needs of the world, but also to solve global environmental problems. Solar energy can be exploited through solar thermal and solar photovoltaic (SPV) routes for various applications. While SPV technology enables direct conversion of sunlight into electricity through semiconductor devices called solar cells, solar thermal technologies utilize the heat energy from the sun for a wide range of purposes.

Solar thermal technologies are quite diverse in terms of their operational characteristics and applications—they include fairly simple technologies such as solar space heating and solar cooking as well as complex and sophisticated ones like solar air conditioning and solar thermal power generation. Solar thermal technologies have also a broad bandwidth in terms of their economical standing. Solar water heating and solar space heating, for example, are very cost effective and are regarded among the most economical renewable energy technologies while high temperature technologies such as solar thermal power generation and solar air conditioning are on the higher economic bandwidth. Solar thermal technologies on the basis of their working temperature can be classified into the following three types:

- Low-temperature technologies (working temperature $<70°C$)—solar space heating, solar pond, solar water heating, and solar crop drying.
- Medium-temperature technologies ($70°C <$ working temperature $<200°C$)—solar distillation, solar cooling, and solar cooking.
- High-temperature technologies (working temperature $>200°C$)—solar thermal power generation technologies such as parabolic trough, solar tower, and parabolic dish.

In the coming sections, this entry provides an overview of these technologies, briefly highlighting their technology fundamental and operating principles.

SOLAR SPACE HEATING

Solar energy can be used to accomplish heat for comfort in buildings. The application, referred to as solar space

Keywords: Solar thermal technologies; Solar radiation; Concentrated collectors; Parabolic trough; Sorption.

heating, is used to optimize the reduction of auxiliary energy consumption in such a way that minimum overall cost is obtained. In combination with conventional heating equipment, solar heating provides the same levels of comfort, temperature stability, and reliability as conventional systems. A building that includes some arrangement to admit, absorb, store, and distribute solar energy as an integral part is also referred as a solar house. A solar space-heating system can consist of a passive system, an active system, or a hybrid of the two.

Passive Space Heating

In passive space heating, the buildings are designed or modified so that they independently capture, store, and distribute solar heat throughout the building without using any electrical or mechanical equipment. Inherently flexible passive solar design principles typically accrue energy benefits with low maintenance risks over the life of the building. The design does not need to be complex, but it does involve knowledge of solar geometry, window technology, and local climate. Passive solar heating techniques generally fall into three categories: direct solar gain, indirect solar gain, and isolated solar gain, as shown in Fig. 1.

Direct Solar Gain Design

Direct gain passive designs typically have large windows with predominately equatorial aspects. In this design, solar radiation directly penetrates and is stored in the building's inherent thermal mass, in materials such as concrete, stone floor slabs, or masonry partitions that hold and slowly release heat.

Indirect Solar Gain Design

In indirect solar passive design, a glazed heat collector, also referred as Trombe wall, collects and stores solar radiation during the day. A Trombe wall consists of an 8–16″ thick masonry wall coated with a dark, heat-absorbing material and covered by a single or double layer of glass, placed from about 3/4″ to 6″ away from the masonry wall. Heat from the sun is stored in the air space between the glass and the dark material, and conducted slowly to the interior of the building through the masonry through the conduction and convection mechanisms.

Isolated Solar Gain Design

In isolated solar passive systems, an extra highly glazed unheated room—a sun-space or conservatory—is added to the south side of the house. Solar gains always make sun-spaces warmer than the outside air, and this reduces heat losses from the house and warms any ventilation air that passes through the sun-space. When solar gains raise the sun-space above house temperature, the heat collected can be let into the house by opening communicating doors and windows.[2]

Active Space Heating

In active space heating of buildings, additional electrical and mechanical equipment is incorporated to circulate solar heated water or air. The main components of an active system are the heat collectors, storage tanks or pebble bed storage, heat exchangers, heat emitters, fans/pumps, connecting pipes or ducts, and controls. Active solar heating systems can be designed to provide the same levels of control of condition in the heated spaces as conventional systems. With indoor temperature essentially fixed at or little above a minimum, load estimations can be done by conventional methods. Passively heated buildings in many cases are not controlled within the same narrow temperature ranges.[3]

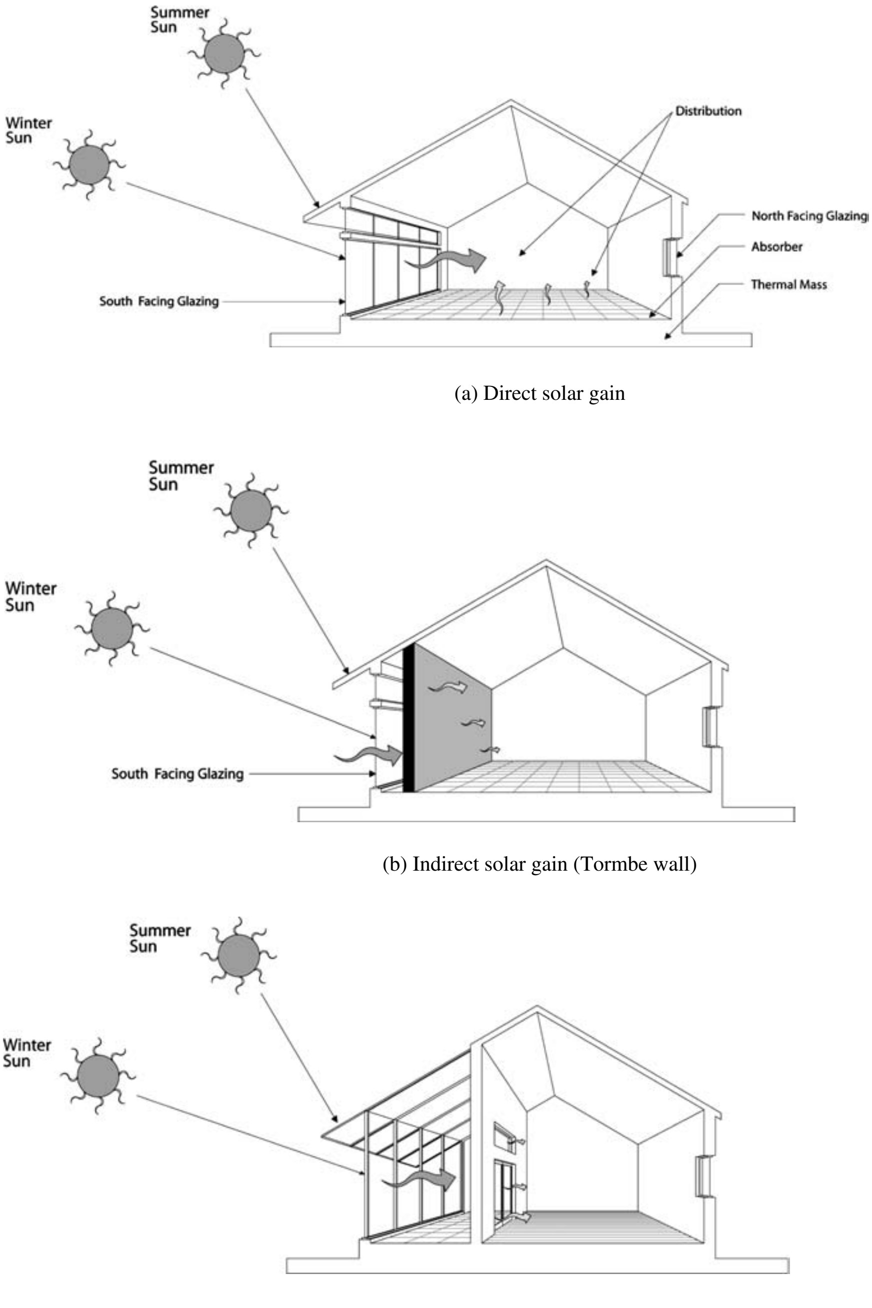
Hybrid Solar Space Heating

Solar space heating system can also be of hybrid nature, combining both the passive and active modes. For example, in a hybrid system, a roof-space collector accomplishes passive collection of solar energy that can be actively distributed in the house using a fan and associated ductwork.

SOLAR WATER HEATING

Water heating is an essential feature of energy requirements in industrial and commercial sectors in general and in domestic sector in particular. A solar water heater consists of two main elements—the collector and the water storage tank, which respectively have the functions of absorbing solar radiation and transferring it to the water, and storing the water for usage. The collectors in solar water heaters can be broadly classified into two categories—flat plate and evacuated tube. A flat plate collector consists of an absorber plate that absorbs solar energy while a glazing above it is used to reduce convective heat loss. An evacuated tube collector consists of tubes with vacuum maintained between the tubes and glazing for better protection against convective heat loss.

Solar water heaters come in three main types: thermosyphon, built-in-storage, and forced circulation. There are two operating principles for solar water heaters: passive system, which relies on natural circulation of water (such as thermosyphon, built-in-storage types); and active system, which uses an external element such as an electric pump to circulate the water (such as the forced circulation type). Another criterion that distinguishes solar water heaters is the



(a) Direct solar gain

(b) Indirect solar gain (Tormbe wall)

(c) Isolated solar gain (sunspace)

Fig. 1 Passive solar space heating principles.

way they transfer heat to water. Again, there are two types: direct system, in which the collector itself transfers heat to water; and indirect system, in which a heat-transfer fluid, circulating in collector in a closed loop, transfers heat to water through a heat exchanger. Fig. 2 shows an indirect active solar water heater.

The efficiency of a solar water heater depends upon its design and the available solar radiation. In this entry, solar water heating has been classified as a low-temperature thermal technology because most of its application is in residential sector where it operates at a temperature of $\leq 70°C$. Also, in industrial applications, solar water heaters are used as preheaters and to hold supply water at almost the same temperature for further heating by conventional means. Solar water heaters are a cost-effective technology—the payback period of solar water heaters can be as low as 3 years while having a service life of more than 20 years. Owing to its technical and economical viability across the world, solar water heating is one of the most established and efficient application of solar energy. Among solar thermal technologies, solar water heating holds the greatest market share and the highest market growth rate.

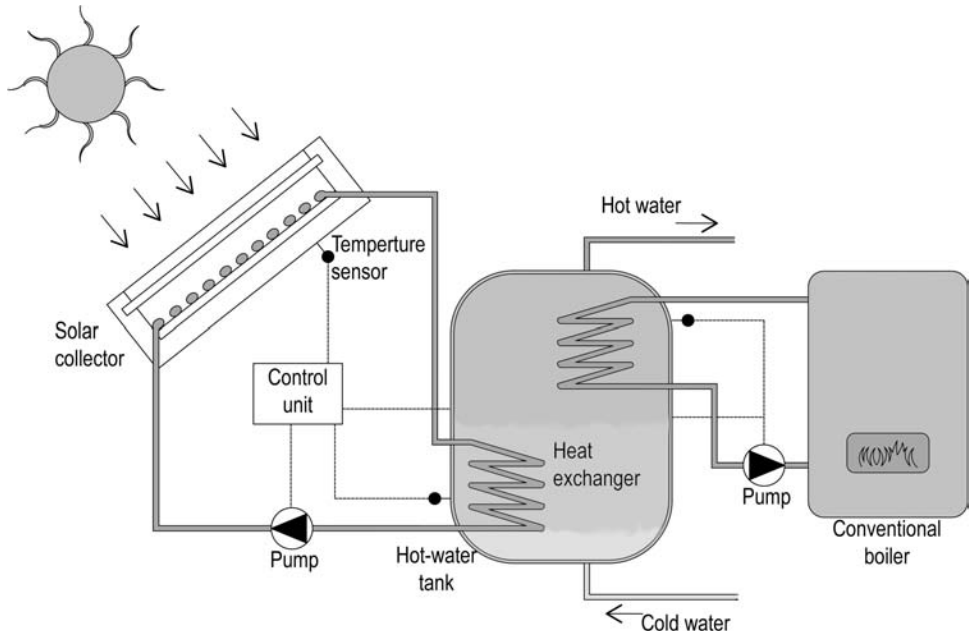


Fig. 2 Indirect active solar water heating system.

SOLAR PONDS

Solar ponds are naturally occurring salt gradient lakes that collect and store solar energy. A solar pond contains salt water with increasing concentrations of salt, hence the density of the solution. When solar radiation is absorbed, the density gradient prevents heat in the lower layers from moving upward by convection and leaving the pond. This results in an increased temperature at the bottom of the pond and a near atmospheric temperature at the top of the pond. The phenomenon of solar ponds was first discovered in 1902 by von Kalecsinsky, who reported that the Medve Lake in Transylvania, containing nearly saturated NaCl solution at a few meters depth with almost fresh water at its surface, had a bottom temperature of 70°C.

A solar pond has three distinctive zones. The top layer is the surface zone that has a low salt content and is at atmospheric temperature. It is also called the upper convective zone (UCZ), as shown in Fig. 3. The bottom layer has a very high salt content and is at a high temperature, 70°C–90°C. This is the zone that collects and stores solar energy in the form of heat and it is called the lower convective zone (LCZ). There is an intermediate insulating zone with a salt gradient. It establishes a density gradient that prevents heat exchange by natural convection, and hence it is called the nonconvective zone (NCZ). In this zone, salt content increases with depth, creating salinity.

Solar ponds can be broadly classified into two main types: nonconvective and convective. In nonconvective solar ponds, the heat loss to environment is reduced by suppressing natural convection normally by using salt stratification. While in convective ponds, heat loss to environment is reduced by covering the pond surface with

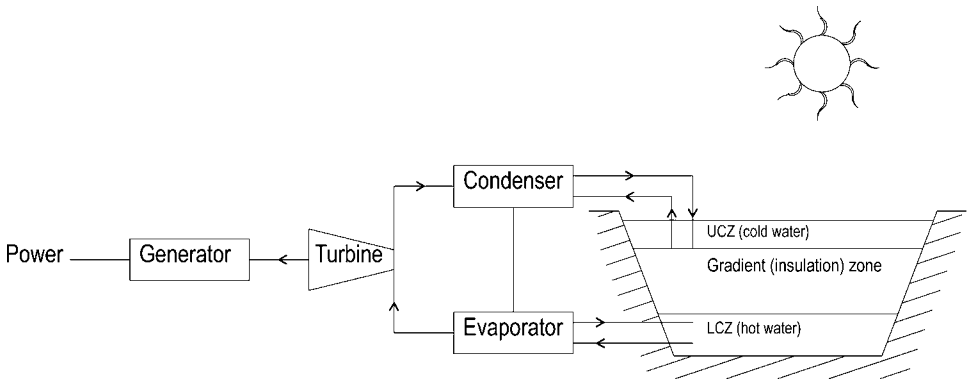


Fig. 3 Schematic of solar pond power generation system.

a transparent material. The heat trapped in the solar ponds can be used for many different purposes, such as industrial process heating, the heating of buildings, desalination, and to drive a turbine for generating electricity.

The first artificial solar pond was developed in Israel in 1958. Since then, many countries such as Australia, the United States, China, India, Iran, Italy, and Mexico have constructed solar ponds, mostly for research and development purposes. During the last decade, significant success in operational practices and applications of solar pond technologies has been achieved.[4]

SOLAR CROP DRYING

Drying is the oldest technique used to preserve food. Until around the end of the 18th century when canning was developed, drying was virtually the only method of food preservation. Solar energy is the main driving force that utilizes warm air to dry food. In drying, the moisture from the food is reduced to a certain level—as low as 5%–25% depending on the type of food-to prevent decay and spoilage in an environment free of contaminations such as dust and insects. Successful drying depends on[5]:

- Enough heat to draw out moisture, without cooking the food
- Dry air to absorb the released moisture
- Adequate air circulation to carry off the moisture

Solar drying can be carried out in open air under the sun by simply spreading the material on a clean surface or in particularly designed solar dryers. Solar dryers, however, exhibit many advantages over open air drying. Firstly, solar dryers are more efficient because they require lesser drying time and area. Secondly, the product is protected from rain, insects, animals, and dust, which may contain faecal material. Thirdly, faster drying reduces the likelihood of mold growth. Fourthly, higher drying temperatures mean that more complete drying is possible, and this may allow much longer storage times (only if rehumidification is prevented in storage). Finally, more complex types of solar driers allow some control over drying rates. Solar dryers can be made in many different designs depending upon various factors, i.e., the type of produce, scale of operation, and local economical and environmental conditions. In terms of their operational mode, solar dryers can be broadly classified into two main types, active and passive dryers, which can both be further subclassified into direct (in which the produce is directly heated from sun) and indirect types (in which the produce is not directly exposed to sun).

Almost all types of food—for example, vegetables, fruits, milk, herbs, spices, meat, and fish—can be dried by solar energy. The advantages of solar food drying are numerous. Dried foods, for example, are tasty, nutritious, lightweight, easy to prepare, and easy to store and use. The energy input is less than what is needed to freeze or can, and the storage space is minimal compared with that needed for canning jars and freezer containers.

SOLAR DISTILLATION

Solar distillation is a process that utilizes solar energy to purify water through evaporation and condensation processes. The process is also referred as water desalination when solar energy is used to purify water from saline water. Solar water distillation is a solar technology with a very long history. Installations were built over 2000 years ago, although they were to produce salt rather than drinking water. Documented use of solar stills (the distillation unit) began in the 16th century. An early large-scale solar still was built in 1872 that spread over an area of 4600 m^2 capable of producing 23,000 liters of drinking water for a mining community in Chile. Mass production occurred for the first time during the World War II when 200,000 inflatable plastic stills were made to be kept in life-crafts for the U.S. Navy.[6] In addition to their use in obtaining drinking water, solar stills are also suitable for the production of distilled water if there is appreciable demand for it in industry, laboratories, and medical facilities or to fill lead acid batteries.

Solar stills come in different designs; however the main features of operation are the same for all of them. In its simple form, water can be placed in an airtight basin that has a sloped transparent cover normally made of glass or plastics, though glass is preferred for its high transparency. The basin is coated with a black lining to maximize absorption of solar radiation. The incident solar radiation is transmitted through the glass cover and is absorbed as heat by the black surface in contact with the water to be distilled. The water is thus heated and gives off water vapor. The vapor condenses on the glass cover, which is at a lower temperature because it is in contact with the ambient air, and runs down into a tray where it is fed to a storage tank, as shown in Fig. 4. The economic viability of solar stills is determined to a critical degree by the design, the construction, the materials employed, and the local market conditions.

SOLAR COOKING

A solar cooker or solar oven harnesses solar energy to cook food. The solar cooker was first developed by a Swiss scientist Horace de Saussure in 1767.[7] Solar cookers are now being used in many countries across the world, especially in remote areas of poor countries. Solar cookers accomplish free cooking with environment friendliness as they only capitalize solar energy. Solar cooking can be very helpful in reducing the deforestation and pollution

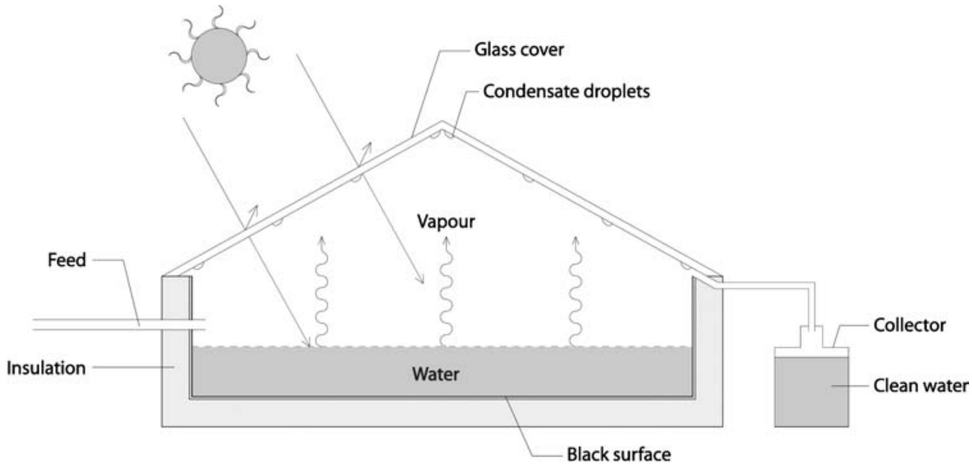


Fig. 4 Schematic diagrams of a solar still.

that originate from consumption of wood, and animal and agricultural residues for cooking in remote areas that lack access to electricity and gas. Solar cookers are capable of performing various types of cooking phenomenas, i.e., frying, baking, and boiling. The maximum achievable temperature depends on the intensity of the available solar radiation and the design and size of the solar cooker. Solar cookers come in a wide range of designs, which can be categorized under the following three major types.

Solar Box Cookers

A solar box cooker consists of an insulated box with a transparent top and a reflective lid. It is designed to capture solar radiation and make use of the greenhouse effect to cause heat to accumulate inside. The top is removable to allow food pots to be placed inside. Temperatures in a typical box cooker can reach above 200°C, but the temperatures achieved obviously depend on the size and design parameters of the cooker and the location of use.

Solar Panel Cookers

The solar panel cooker is the simplest solar cooker, and it consists of multiple simple reflectors arranged to focus solar radiation onto a covered black pot enclosed in a clear heat-resistant plastic bag or other transparent enclosure, such as glass bowel.

Solar Parabolic Cookers

Parabolic solar cookers, also called concentrated cookers, consist of a concave disk that focuses the light onto the bottom of a pot that is arranged at the focal length of the disk, as shown in Fig. 5. These are the most efficient types of solar cookers.

SOLAR COOLING/AIR CONDITIONING

Solar thermal energy can be used for cooling and dehumidification. Collectors play a critical role in extracting the energy from solar radiation to operate the cooling device. The collectors used in solar thermal cooling could be of various types, such as low-temperature

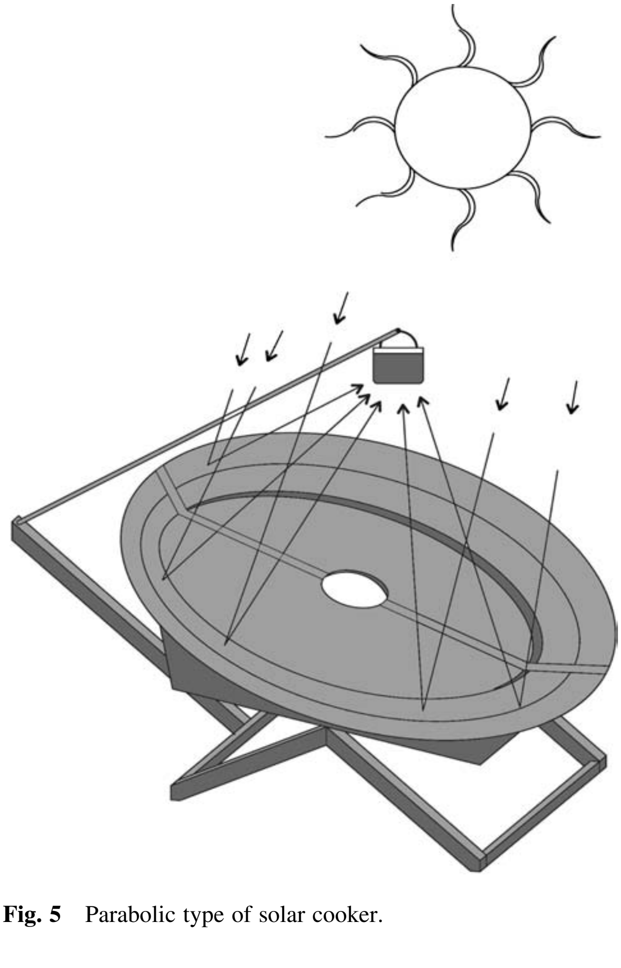


Fig. 5 Parabolic type of solar cooker.

flat plates and high-temperature evacuated tubes and concentrators. The basic principle behind solar thermal cooling is the thermochemical process of sorption—a liquid or gaseous substance is either attached to a solid, porous material (adsorption) or it is taken in by a liquid or solid material (absorption). The heat transfer fluid is heated in the solar collectors to a temperature well above ambient and used to power a cooling device—a type of heat-actuated pump. The heat transfer fluid may be air, water, or another fluid; it can also be stored in a hot state for use during times of no sunshine. Heat extracted by the cooling device from the conditioned space and from the solar energy source is rejected to the environment using ambient air or water from a cooling tower.[8]

The solar thermal cooling process can be broadly classified under open cycle systems and closed cycle systems. Open cycle systems are those in which the refrigerant is in direct contact with atmosphere and is discarded from the system after providing the cooling effect and new refrigerant is supplied in an open-ended loop. In closed systems, on the other hand, the refrigerant is not in direct contact with the atmospheric air. Open and closed cycle systems can further be distinguished according to the type of sorbent used, which can be in a liquid or a solid form. The three main designs of solar thermal cooling technologies that have gained the most attraction include solar adsorption, solar absorption, and solar desiccant. The key features of these designs are provided in Table 1.[9]

SOLAR THERMAL POWER GENERATION

Solar thermal power generation systems start with capturing heat from solar radiation. Direct solar radiation can be concentrated and collected by a range of concentrating solar power technologies to provide medium- to high-temperature heat. This heat then operates a conventional power cycle—for example, through a steam turbine or a Stirling engine to generate electricity. Solar thermal power plants can be designed for solar-only or hybrid operation, where some fossil fuel is used in case of lower radiation intensity to secure reliable peak-load supply. Five distinct solar thermal power generation concepts are available:

- Solar pond
- Solar chimney
- Solar parabolic trough
- Solar central receiver or solar tower
- Solar parabolic dish

Solar pond and solar chimney are nonconcentrated types of technology. In this section, the three concentrated types of solar thermal technologies—solar parabolic trough, solar central receiver, and solar parabolic dish—are discussed, as they have received the greater degree of attention over the years due to their favorable technical and commercial characteristics. These technologies can be used to generate electricity for a variety of applications, ranging from remote power systems as small as a few kilowatts (kW) up to grid-connected applications of 200–350 megawatts (MW) or more.[10]

Solar thermal power generation systems have three essential elements needed to produce electricity: a concentrator (to collect and focus solar radiation), a receiver (to convert concentrated solar radiation into heat), and an engine cycle (to generate electricity). Some systems also involve a transport or storage system. Solar collectors have a crucial role to play in the whole system and can be mainly classified into two types: concentrating and nonconcentrating. They are further categorized on the

Table 1 Overview of processes for thermally powered cooling and air conditioning

Solar thermal cooling system design	Adsorption refrigeration	Absorption refrigeration	Desiccant air conditioning
Solar collector	Vacuum tube collector, flat plate collector	Vacuum tube collector	Flat plate collector, solar air collector
Coolant circulation process	Closed refrigerant circulation systems	Closed refrigerant circulation systems	Open refrigerant circulation systems (in contact with the atmosphere)
Process basic principle	Cold water production	Cold water production	Air dehumidification and evaporative cooling
Sorbent type	Solid	Liquid	Solid
(Refrigerant/sorbent)	Water–silica gel ammonia–salt	Water–water–lithium bromide, ammonia–water	Water–silica gel water–lithium chloride–cellulose
Typical operating temp.	60°C–95°C	80°C–110°C (one step) 130°C–160°C (two step)	45°C–95°C

(a) Parabolic trough

(b) Central receiver or solar tower

(c) Parabolic dish (Stirling engine)

Fig. 6 Solar thermal power generation technologies.

basis of their concentrator optical properties and the operating temperature that can be obtained at the receiver. Most of the techniques for generating electricity from heat need high temperatures to achieve reasonable efficiencies. Concentrating systems are hence used to produce higher temperatures. Table 2 shows the operational characteristics of concentrated collectors.[11]

Parabolic Trough

The parabolic trough systems consist of large curved mirrors or troughs that concentrate sunlight by a factor of 80 or more onto thermally efficient receiver tubes placed in the trough's focal line, as shown in Fig. 6a. A thermal transfer fluid, such as synthetic thermal oil, is circulated in the tubes at focal length. Heated to approximately 400°C by the concentrated sun's rays, this oil is then pumped through a series of heat exchangers to produce superheated steam.[12] The steam is converted to electrical energy in a conventional steam turbine generator, which can either be part of a conventional steam cycle or integrated into a combined steam and gas turbine cycle, as shown in Fig. 7. Parabolic trough power plants are the only type of solar thermal power plant technology with existing commercial operating systems.

It is also possible to produce superheated steam directly using solar collectors. This makes the thermal oil unnecessary and also reduces costs because the relatively expensive thermo oil and the heat exchangers are no longer needed. However, direct solar steam generation is still in the prototype stage.

Central Receiver or Solar Tower

In solar thermal tower power plants, hundreds or even thousands of heliostats (large individually tracking mirrors) are used to concentrate sunlight onto a central receiver mounted at the top of a tower, as indicated in Fig. 6b. A heat-transfer medium in this central receiver absorbs the highly concentrated radiation reflected by the heliostats and converts it into thermal energy to be used for the subsequent generation of superheated steam for turbine operation. To date, the heat transfer media demonstrated include water or steam, molten salts, liquid sodium, and air. If pressurized gas or air is used at very high temperatures of about 1000°C or more as the heat transfer medium, it can even be used to directly replace natural gas in a gas turbine, thus making use of the excellent cycle of modern gas and steam combined cycles.

Parabolic Dish

A parabolic dish system uses a parabolic concave mirror to concentrate sunlight onto a receiver located at the focal point of the mirror, as highlighted in Fig. 6c. The concentrated beam radiation is absorbed into the receiver to heat a fluid or gas (air) to approximately 750°C. This fluid or gas is then used to generate electricity in a small piston, Stirling engine, or a microturbine attached to the receiver. These systems stand alone, and they are normally used to generate electricity in the kilowatts range.[13]

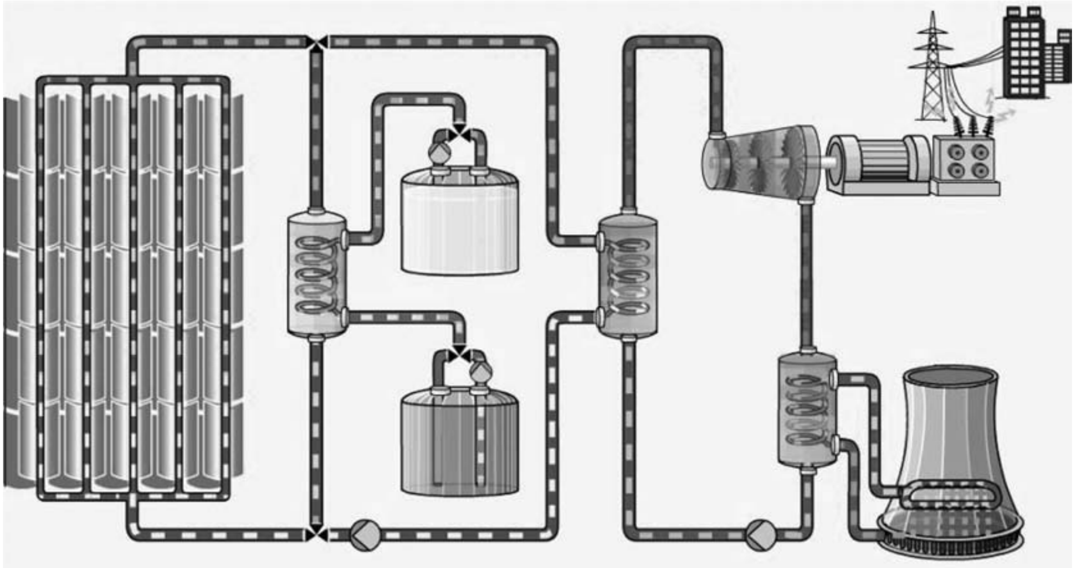


Fig. 7 Schematic of solar parabolic trough power plant.

SOLAR THERMAL TECHNOLOGIES—MARKET GROWTH AND TRENDS

Solar thermal technologies, in general, like SPV and other renewable energy technologies, are rapidly growing. The two key growth areas in solar thermal technologies, however, are solar water heating and solar thermal power generation. Solar water heating is among the fastest growing renewable technologies. As it was reported in the year 2003, solar water heaters received a 21% share of the total investment, $22 billion, in the renewable energy sector worldwide.[14] In 2004, China, which presently holds almost 80% of the installed solar thermal collector area, found its market growth by 30% as it installed well over 10 million m^2 of new solar thermal collector areas—equivalent to 7 GWth. There are similar growth trends across the world, especially in European countries.

The solar thermal power market has experienced a relative state of stagnancy since the early 1990s. However, new opportunities are now opening up for solar thermal power as a result of the global search for clean energy solutions. Both national and international initiatives are supporting the technology, encouraging commercialization of production. Recently, commercial plans in Spain and the United States have led a resurgence of interest, technology evolution, and potential investment. Some developing countries, including Morocco, India, Egypt, and Mexico have also planned projects with multilateral assistance. Examples of specific large solar thermal projects currently under construction or in advanced permitting and development stage around the world include:

- *Spain*. Over 500 MW solar capacity using steam cycle (4×10–20 MW solar tower and 12×50 MW parabolic trough)
- *Morocco*. 220-MW Integrated Solar Combined Cycle (ISCC) plant with 30 MW solar capacity (trough)
- *United States*. 50-MW solar capacity with parabolic trough in Nevada using steam cycle, preceded by a 1-MW parabolic trough demonstration plant using Organic Rankine Cycle (ORC) turbine in Arizona
- *United States*. 500-MW Solar Dish Park in California, preceded by a 1-MW (40×25 kW) test and demo installation

Table 2 Characteristics of typical concentrated solar collectors

Solar collector technology	Typical operating temperature (°C)	Concentration ratio	Tracking	Maximum conversion efficiency (Carnot) (%)
Solar fresnel reflector technology	260–400	8–80	One-axis	56
Parabolic trough collectors	260–400	8–80	One-axis	56
Heliostat field + central receiver	500–800	600–1000	Two-axis	73
Parabolodial dish concentrators	500–1200	800–8000	Two-axis	80

- *Italy*. 40 MW solar capacity integrated into existing combined cycle plant (trough)
- *Mexico*. 291-MW ISCC plant with 30 MW solar capacity (trough)
- *Algeria*. 140–150 MW ISCC plant with 25 MW solar capacity (trough)

A scenario of what could be achieved by the year 2025 was prepared by Greenpeace International, the European Solar Thermal Industry Association, and International Energy Agency (IEA) SolarPACES projects. From the current level of just 354 MW total installed capacity, the rate of annual installation by 2015 will have reached 970 MW, thus reaching a total installed capacity of 6454 MW. By 2025, 4600 MW will come on stream each year. According to this projection, by 2025, the total installed capacity of solar thermal power around the world will reach over 36,000 MW. It is also projected that by 2040 more than 5% of the world's electricity demand may be satisfied by solar thermal power.[12]

CONCLUSIONS

Solar thermal technologies operate by converting solar radiation into heat, which can be either directly utilized in various applications such as solar space heating, solar water heating, and solar air conditioning, or can be transformed into electricity to serve any purpose similar to conventional electricity. The key element in all solar thermal technologies is the collector, whose function is to gather the heat of solar radiation. Collectors normally come in three different types: flat plate, evacuated tube, and concentrated, and they operate in a wide range of temperatures, i.e., from less than 50°C to more than 1200°C. Solar thermal technologies normally operate in passive or active modes. Different types of solar thermal technologies are gaining huge attention across the world depending upon their technical and economical viability. Solar water heating, for example, in 2003, received a 21% share of the total investment made in renewable energy sector worldwide. Solar thermal power generation is also expected to grow at a healthy rate in coming years, as it is projected that by 2040 more than 5% of the world's electricity demand could be satisfied by solar thermal power.

REFERENCES

1. http://www.mtpc.org/index.asp (accessed on March 2006).
2. Muneer, T. Solar energy. *KEMPS Engineering Year Book*, Miller Freeman: New York, 2000.
3. Duffie, J.; Beckman, W. *Solar Engineering of Thermal Processes*, 2nd Ed.; Wiley: New York, 1991.
4. Akbarzadeh, A.; Andrews, J.; Golding, P. Solar pond technology: A review and future directions. *Advances in Solar Energy*, Earthscan: London, 2005.
5. Whitfield, D.E. *Solar Drying*, International Conference on Solar Cooking, South Africa, November 26–29, 2000.
6. Solar distillation, Technical Brief, Intermediate Technology Development Group.http://www.itdg.org/?id=technical_briefs (accessed on March 2006).
7. http://en.wikipedia.org/wiki/Solar_oven (accessed on March 2006).
8. Grossman, G. Solar cooling, dehumidification, and air-conditioning, solar thermal power generation. In *Encyclopedia of Energy*, Vol. 5; Elsevier: Amsterdam, U.K., 2004.
9. http://www.solarserver.de/solarmagazin/artikeljuni2002-e.html (accessed on March 2006).
10. Concentrating Solar Power: Energy from Mirrors, report Produced by NREL 2001. Available at: http://www.nrel.gov/docs/fy01osti/28751.pdf.
11. Luzzi, A.; Lovegrove, K. Solar thermal power generation. In *Encyclopedia of Energy*, Vol. 5; Elsevier: Amsterdam, U.K., 2004.
12. Concentrated solar thermal power—now!, Report by European Solar Thermal Industry Association and Greenpeace, 2005 http://www.greenpeace.org/raw/content/international/press/reports/Concentrated-Solar-Thermal-Power.pdf.
13. Quaschning, V. Solar thermal power plants. Renewable Energy World **2003**, *June*.
14. Annual investment in renewable energy: 1995–2003. Renewable Energy World **2005**, *February*.

Solar Water Heating: Domestic and Industrial Applications

M. Asif
School of the Built and Natural Environment, Glasgow Caledonian University, Glasgow, Scotland, U.K.

T. Muneer
School of Engineering, Napier University, Edinburgh, Scotland, U.K.

Abstract
Solar water heating is one of the most successful applications of solar thermal technologies. It provides environmentally clean energy and has enormous potential within domestic and industrial sectors. Solar water heaters are categorized into three main types: thermosyphon, built-in-storage, and forced circulation. This entry provides the fundamentals of solar water heating technology and a description of parameters that determine the performance of a solar water heating system. Economics of solar water heating and the market trends around the world have also been discussed. Results from a specific life cycle assessment (LCA) undertaken on built-in-storage solar water heater are also presented.

INTRODUCTION

Energy is imperative for human life. The accomplishments of civilization have largely been achieved through the increasingly efficient and extensive harnessing of various forms of energy to extend human capabilities and ingenuity. One of the biggest challenges to mankind in the 21st century is to develop methods of generating and using energy that could meet the rapidly growing energy demands while protecting the global ecosystem. Renewable energy resources such as solar energy, wind power, biomass, and geothermal energy are abundant, inexhaustible, and environmentally friendly. Solar energy is one of the most promising types of renewable energy that has the potential to meet the energy demand of the entire planet. Solar water heating, one of the oldest and the most successful applications of solar thermal technologies, utilizes solar energy to heat water without producing any harmful emissions into the environment.

Solar water heating, besides its domestic role, has a wide array of applications within the commercial sector (e.g., swimming pools, laundries, hotels, and restaurants) and the industrial sector (e.g., food and beverages, processing, and textile industries). Around the world, water heating accounts for as much as 15%–25% of the total energy consumed in the domestic sector. In the United States and United Kingdom, for example, water heating, respectively, consumes 18 and 23% of the domestic energy.[1,2] In the industrial sector, water heating may account for a significantly higher share of energy usage. In the textile sector, for example, water heating can account for as much as 65% of the total energy used during processes such as dyeing, finishing, drying, and curing.[3] Solar water heating systems can also be used for large industrial loads and for providing energy to district heating networks.

A solar water heating system essentially consists of a collector and a water storage tank. The collector absorbs solar radiation and transfers it to the water stored in the tank. Residential, commercial, and a considerable number of industrial applications often require hot water that is at a temperature of less than 60°C. The required hot water temperature, however, could vary depending upon the type of activity especially within the textile sector where it often needs to be near the boiling level. Modern solar water heaters can accomplish these temperatures. However, taking into account various factors such as unpredictable weather conditions (rain or overcast sky) and a solar heating system's size limitation in case of high demand, solar water heaters require a conventional heating system as a back up. A typical domestic solar water heating system can provide up to two-thirds of the total hot water requirements cutting down on fossil-fuel energy costs and also reducing the associated environmental impacts. There is a significant environmental penalty associated with the combustion of hydrocarbon gas with 14,000 g of CO_2 and 65 g of NO_x emitted per GJ of energy released. This can be reduced by using solar water heater.

This entry provides fundamentals of technology for solar water heating and a description of parameters that determine the performance of a solar water heating system. Economics of solar water heating and the market trends around the world are also discussed. Results of a life cycle assessment (LCA) study of a built-in-storage solar water heater are also presented.

Keywords: Solar energy; Solar water heating; Flat plate collectors; Evacuated tube collectors; Renewable energy; Life cycle assessment.

SOLAR WATER HEATER DESIGNS

Solar water heaters can be categorized into three main types: thermosyphon, built-in-storage, and forced-circulation. Thermosyphon and built-in-storage types are also regarded as passive systems as they rely on the natural circulation of water. The forced-circulation type of heater, on the other hand, is regarded as an active system as it incorporates an external element such as an electric pump to circulate the water. Solar water heaters transfer heat to the water in two ways: a direct system in which the collector itself transfers the heat to water, and an indirect system in which a heat-transfer fluid, circulating in the collector in a closed loop, transfers the heat to water through a heat exchanger.

Thermosyphon Solar Water Heater

The thermosyphon system operates on the principle that cold water has a higher density than warm water, and so being heavier, will sink down. Therefore, the collector is always mounted below the water storage tank, so that cold water from the tank reaches the collector via a descending water pipe (see Fig. 1). The collector absorbs solar radiation and transforms it into thermal energy and induces that energy to the water inside. As the temperature of the water in the collector rises, its density decreases. A circulation is thus established which enables the tank water to be progressively heated. The temperature of water at any point in the circulation determines its corresponding density. If this density variation is plotted against the height contour of the circuit, the magnitude of the driving force for water circulation is obtained. Owing to the low flow rates encountered within such systems, the driving force is balanced against the friction offered by the fluid circuit. The flow rates can thus be estimated. A thermosyphon system's storage tank must be positioned well above the collector, since otherwise the cycle can run backwards during the night, cooling down all the heated water. Furthermore, the cycle does not work properly at very small height differences. To avoid flow reversals, it is recommended that the bottom of the tank should be at least 300 mm above the top of the collector. This requirement makes the system inconvenient, particularly if the collectors are to be roof-mounted. Freezing may occur within the collectors during winter in a direct system. The solution is to employ a heat exchanger within the storage tank, with an anti-freeze solution added to water. This will, however, result in reduced system efficiencies and increased system cost.

Built-in-Storage Solar Water Heater

A "built-in-storage water heater," also referred as an "Integral collector-storage system", combines a flat-plate collector and storage tank in one unit. Built-in-storage water heaters can be further classified into plain and finned types.

Plain Built-in-Storage Water Heater

The construction of a built-in-storage water heater involves a rectangular box-like structure with the top face painted black and enclosed behind a single or double sheet of glass. The back surface and sides are properly insulated and the entire assembly is tilted at a suitable inclination. The built-in-storage water heater possesses several advantages over the other types. First, during the daytime, it operates at a higher efficiency owing to the fact that primarily no heat losses occur during water circulation. Second, in comparison with the thermosyphon and forced-circulation systems, the intimate contact of water with the absorber plate results in better heat transfer. Built-in-storage heaters are also compact in size and

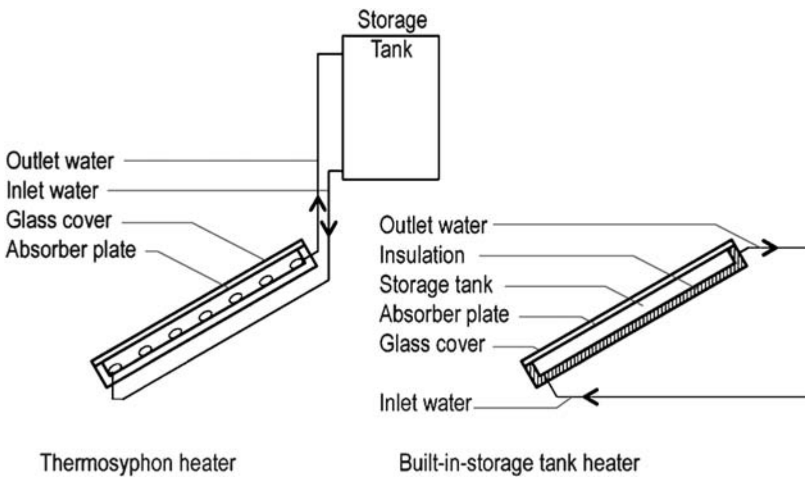


Fig. 1 Schematic diagrams of thermosyphon and built-in-storage solar water heaters.

cheaper due to their simplicity of construction. Based on extensive measurements of a number of designs, Muneer has reported the effects of storage volume/collector area ratio, the number of glazings, and the mode of operation on the performance of built-in-storage heaters.[4] Kreider and Kreith have presented design details of a freeze-tolerant, built-in-storage solar water heater with embedded heat exchanger coil which uses a refrigerant to extract heat.[5] Also in this respect, Davidson and Hammonds have recently undertaken more work.[6,7]

Finned Built-in-Storage Water Heater

The finned type of built in storage solar water heater differs from the plain type only in terms of fins that are incorporated within the thermal collector plate. Fins in this design play a dual role: firstly, they act as a support for the top absorber plate, and thus avoid the bulging due to hydrostatic pressures exerted by the stored water within the heater. Secondly, the fin attempts to enhance the heat transfer process from the absorber plate to the innermost layers of water. Note that within the plain heater the fact that heat is being transferred from a heated plate at the top to a cooler body of water residing underneath is an inefficient convection process. On the other hand, the vertically placed fins have a better opportunity to transfer heat as shown in Fig. 2.[8]

Forced Circulation Solar Water Heater

In a forced-circulation solar water heater, water is actively pumped from the storage tank through the collectors and back into the tank. An electronic controller, a small pump, valves, and other components are needed for proper operation and maintenance. With a forced-circulation water heater, the collector and storage tank can be installed independently, and no height difference between tank and collector is necessary. Two temperature sensors monitor the temperatures in the solar collector and the storage tank. If the collector temperature is higher than the tank temperature by a certain amount, the control starts the pump, which moves the heat transfer fluid in the solar cycle; "switch-on" temperature differences are normally between 5 and 10°C. If the temperature difference decreases below a second threshold, the control switches off the pump again. Fig. 3 shows the schematic of a forced-circulation water heating system. In this design, the problem of placing the solar water tank above the collectors is removed. The water in tank A is the feed water for tank B which has a built-in auxiliary heater. In a more compact version of the forced-circulation system, the two tanks A and B are combined in a single unit. However, this reduces the overall efficiency of the system as the maximum exploitation of solar energy is inhibited.

TYPES OF THERMAL COLLECTORS

The solar thermal collector is at the heart of a solar water heater as it absorbs the solar radiations and then transfers the captured heat to water in the storage tank. Solar thermal collectors may broadly be classified as flat-plate or evacuated-tube collectors.

Flat-Plate Solar Collector

The main components of a flat-plate solar collector include a transparent front cover, a collector housing,

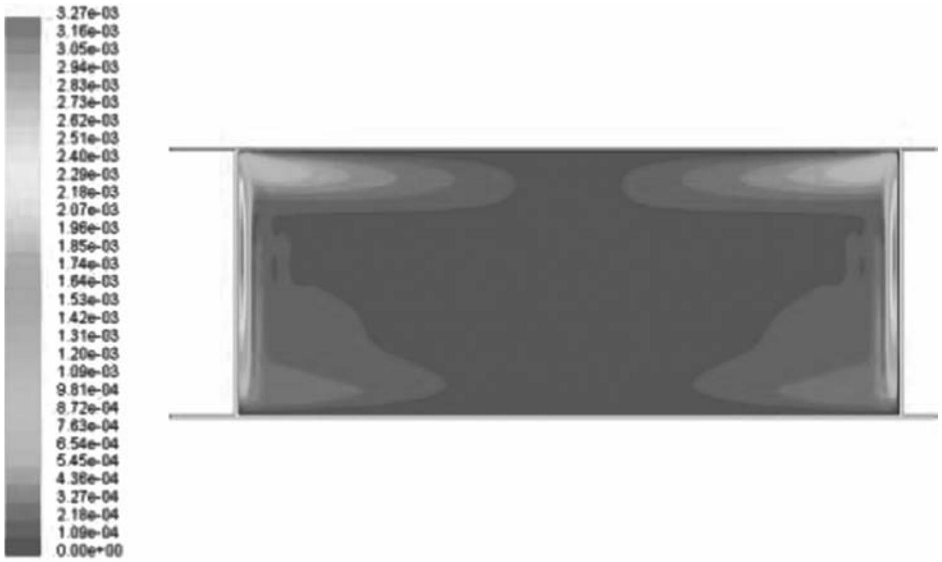


Fig. 2 Computational fluid dynamics (CFD) velocity distribution raster plot for a finned heater. The jets issuing from the left and right fins demonstrate their effectiveness in heating the otherwise quiescent body of water shown.

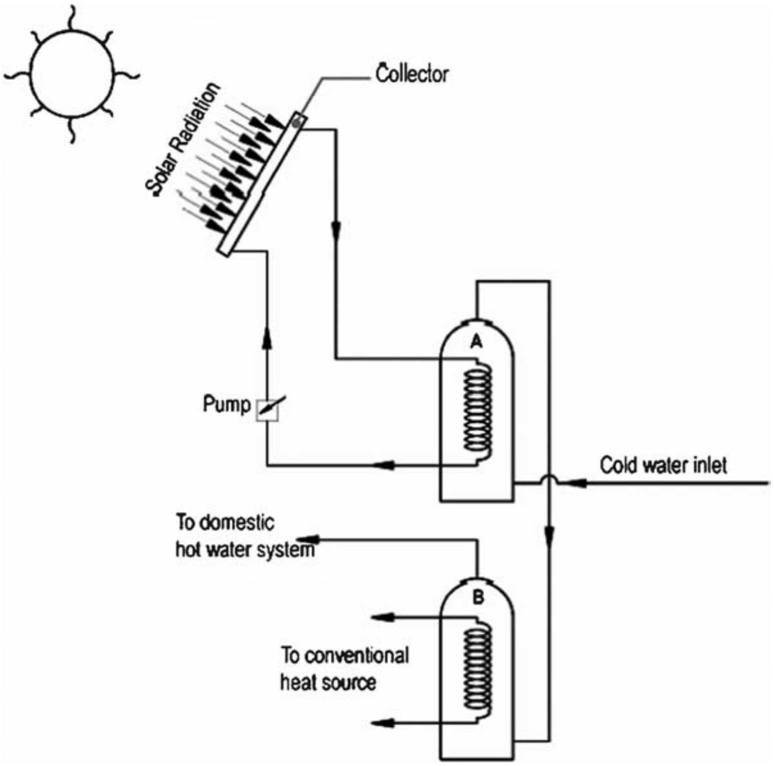


Fig. 3 Forced-circulation solar water heater with indirect cycle.

and an absorber. The absorber is usually made of metal materials such as aluminum, copper, or steel. The collector housing can be made of plastic, metal or wood, and the glass front cover must be sealed so that heat does not escape, and dirt, insects, and humidity do not get into the collector itself. The collector housing is highly insulated at the back and sides, keeping heat losses low. However, there are still minor heat losses from the collector, mainly due to the temperature difference between the absorber and ambient air, and these are subdivided into convection and radiation losses. The former are caused by air movements, while the latter are the product of the exchange of heat by radiation between the absorber and the environment. A sheet of glass covers the collector as it faces the sun, and this helps to prevent most of the convection losses. Furthermore, it reduces heat radiation from the absorber into the environment in a similar way that a greenhouse does. However, the glass also reflects a small part of the sunlight, which does not then reach the absorber at all.

The heat-loss coefficient from any given collector determines its thermal efficiency. This value, in turn, is obtained from the heat loss from the collector top (U_T) and the overall loss coefficient (U_L), i.e., the sum of heat loss from the top, sides, and bottom of the collector housing. For one commercially available single-glazed collector with a selectively coated absorber, U_T and U_L are 3.37 and 3.4 W/m^2-K, respectively.[9]

Eq. 1 is a simplified model for obtaining U_T

$$U_T = \left[\frac{N}{C/T_{pm}[T_{pm} - T_a/(N+f)]^e} + \frac{1}{h_w}\right]^{-1} + \frac{\sigma(T_{pm}+T_a)(T_{pm}^2+T_a^2)}{(\varepsilon_p + 0.00591Nh_w)^{-1} + 2N + f - 1 + 0.133\varepsilon_p/\varepsilon_g - N} \quad (1)$$

Evacuated Tube Solar Collector

Evacuated tube solar collectors normally come in two types, coaxial-tube and heat-pipe designs. In the former design, a coaxial heat exchange pipe carrying a counter-flowing fluid is embedded within the absorber. The heat exchange pipe then feeds into a header. To enable maximum exploitation of solar energy, each vacuum tube is pivot mounted, enabling for the optimal orientation. The vacuum in the glass tubes ensures minimal convective heat loss while the selective coating helps suppress radiative losses. The top-loss and overall-loss coefficients for the collector presently under discussion are quoted as 1.37 and 1.375 W/m^2-K.[9]

The design of the collector shown in Fig. 4 is based on the heat pipe principle. Basically a heat pipe is a closed container that employs an evaporating-condensing cycle.

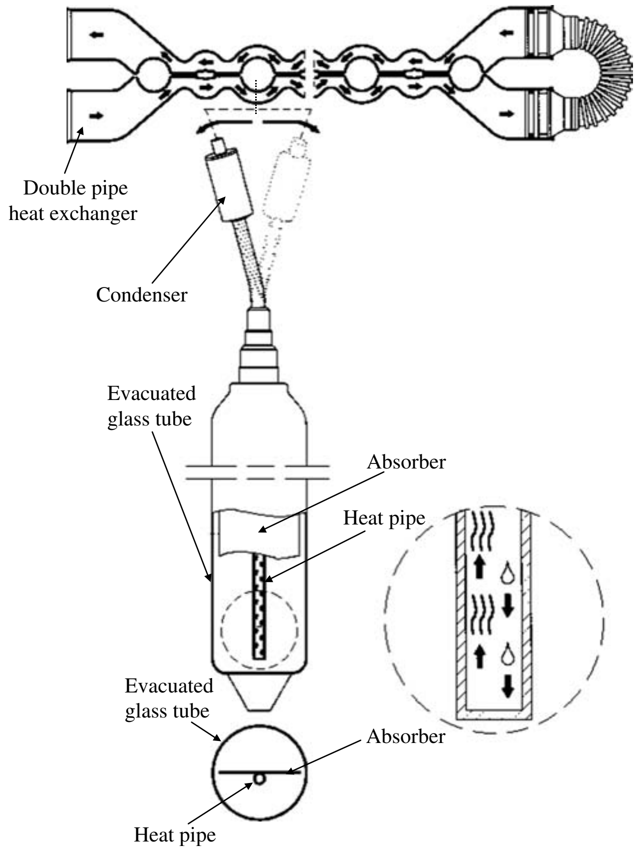


Fig. 4 Evacuated tube solar collector design based on the heat pipe principle.

The transfer of heat from the absorber plate occurs via an efficient and fast heat conductor bearing a low heat capacity. The heat-pipe functions like a thermal diode. It accepts heat from an external source, uses this heat to evaporate the liquid (latent heat) and then releases latent heat by reverse transformation (condensation) to the heat sink. This process is repeated continuously by a return feed mechanism of the condensed fluid back to the heat zone. Typically, evacuated heat pipe solar collectors use a sealed steel pipe, containing alcohol. The pipe is then attached to a black copper fin that fills the tube (absorber plate). Protruding from the top of each tube is a metal tip attached to the sealed pipe (condenser). These tubes feed into a heat exchanger manifold. Under solar flux, the alcohol is heated and hot vapor rises to the top of the pipe. Water or glycol flows through the manifold and picks up the heat from the tubes. The heated liquid circulates through another heat exchanger and gives off its heat to water that is stored in a solar storage tank. The maximum operating temperature of a heat pipe is the critical temperature of this heat transfer medium. Since no evaporation/condensation above the critical temperature is possible, the thermodynamic cycle is interrupted when the temperature of the evaporator exceeds the critical temperature. The top-loss and overall-loss coefficients for the above collector are quoted as 1.34 and 1.346 W/m^2-K.[9]

SYSTEM DESIGN

The rule of thumb for a solar water heating system storage design is to use an 80 l/m^2 collector area for tropical locations, 60 l/m^2 for the United States, Central Europe, and Canada, 40–45 l/m^2 for the British Isles and Scandinavian countries.[1] Losses in the generation of domestic hot water maybe summarized as draw-off pipe losses, storage losses, losses in primary circulation pipes and those encountered during the heating up of the hot water generator. For a realistic plant simulation, these losses should be taken into account. In this section, basic mathematical models are presented which enable system simulation and design evaluation of solar water heating systems.

The Hottel–Whillier–Bliss Equation

An important relationship, which essentially represents the steady-state energy balance of any solar heat collector with an absorber area of A_c, is the Hottel–Whillier–Bliss equation (Eq. 2). This equation enables the calculation of useful energy gain (Q_u) as a function of the fluid inlet temperature, T_i and ambient temperature T_a.

$$Q_u = A_c F_R [I_{G,TLT}(\tau\alpha) - U_L(T_i - T_a)] \tag{2}$$

Eq. 1 enables the computation of U_T. It is customary to add 5%–10% to the value of U_T to obtain the total heat loss coefficient, U_L. Other terms used in Eq. 2 are explained below.

The Transmission-Absorption Product, $(\tau\alpha)$

A solar radiation incident on any inclined collector is composed of direct (beam), sky-diffuse, and ground-reflected components. These transmitted components, passing through the collector glazing, may be estimated if the corresponding transmittances τ_b, τ_d, and τ_g are known. Each of these is a function of their respective irradiance incidence angles. Furthermore, part of the transmitted radiation is reflected back from the absorber plate to the cover system, which is, in turn, absorbed and reflected back to the collector plate. Therefore, as Duffie and Beckman have suggested, the transmittance-absorbance product $(\tau\alpha)$ should be thought of as a symbol representing a property rather than a straightforward product of the two fundamental properties, τ and α. Duffie and Beckman[10] have presented a graphical solution for obtaining $(\tau\alpha)$ from $(\tau\alpha)_n$.

Fin Efficiency (F')

The fin efficiency factor, F' represents the ratio of the actual energy gain to the energy gain that would result if the entire absorbing surface was maintained at the fluid temperature. Duffie and Beckman have presented the physical model for obtaining F' which is shown to be a strong function of the absorber fin heat transfer efficiency, U_L, and other less significant parameters.[10] For most practical designs, F' may be taken to have a value between 0.8 and 0.9.

Heat Removal Factor (F_R)

The collector heat removal factor, F_R, is the product of F' and the flow factor, F''. F_R is equivalent to the effectiveness of the solar collector's heat exchange process. The flow factor is obtained from Eq. 3, which, in turn, enables the estimation of F_R.

$$F'' = (mC_p/A_c U_L F')[1 - \exp(-mC_p/A_c U_L F')] \quad (3)$$

Measured Collector Performance

The basic method of obtaining collector performance is to operate the collector in a steady-state mode with coincident measurements of solar radiation, fluid flow rate, temperature gain, ambient temperature, and wind speed. In Eq. 2 the useful solar heat gain was written in terms of collector parameters. The efficiency of the collector is defined as

$$\eta = Q_u/I_{G,TLT} \quad (4)$$

Eqs. 2 and 4 may be manipulated to express the efficiency as a linear function of the ratio of the temperature differential between the collector plate and the ambient air and the collector irradiation. Thus,

$$\eta = F_R(\tau\alpha) - F_R U_L[(T_i - T_a)/I_{G,TLT}] \quad (5)$$

The European practice is to base the collector test results on $T_{p,av}$ which is the average of the inlet–outlet fluid temperature. Therefore, in Eq. 5 T_i is replaced with $T_{p,av}$.

Eq. 6 may be simplified as

$$\eta = a - bX \quad (6)$$

The X variable represents the term contained within the square brackets of Eq. 5.

Contrary to the dictum of Eq. 5, measured data show a slight non-linear variation of the data points, and as such, second-order models have been suggested for the above efficiency variation. For most practical purposes, though, a straight line fit should suffice.

System Simulation

Eqs. 1–5 enable a fairly representative, hourly simulation of solar water heating plants to be undertaken. A full-blown simulation is beyond the scope of this entry. A detailed simulation exercise would require collector thermal capacitance and tank temperature stratification effects to be taken into account. Hourly (and where applicable, temperature-dependent) estimation of F_R, $(\tau\alpha)$, and U_L may be undertaken with the hot water draw-off profile also incorporated within the simulation program.

ECONOMICS OF SOLAR WATER HEATING

Solar water heaters provide energy with economy. The economic payback period of a solar water heating system, however, is a function of various factors such as system efficiency, local weather conditions (i.e., the level of solar irradiance available), and the cost of fossil fuels. The initial cost of a solar water heater is higher than that of a gas water heater or an electric water heater that varies from region to region. A solar water heater can, however, be much more economical over the lifetime of the system than heating water with electricity, fuel oil, propane, or natural gas, depending upon the price of fuel sources. According to the findings of the Florida Solar Energy Center, the payback period for a well-designed and properly installed solar water heater could be between 4 and 8 years.[11] After the payback period, savings can be accrued over the life of the system, which ranges from

Table 1 Life cycle assessment of a built-in-storage solar water heater

Entity	Quantity (kg)	Embodied energy (MJ)	Carbon released (kg)	Monetary costs (USD)
Stainless steel	33	1155	20.13	108
Glass	11	340	6.70	5
Glass wool	2	40	1.75	2
Rubber	0.1	15	0.28	1
Timber	20	44	0.8	4
Total		1744	39.7	140
Savings/year		3509	69.6	20
Payback period		166 days	156 days	6.1

15 to 40 years, depending on the system and how well it is maintained. Muneer and Asif estimated the payback period for solar water heating incorporated within textile industries in Pakistan to be 6 years. More recent work has concluded that, with more efficient designs, the payback period can be reduced to just over 3 years.[8]

LIFE CYCLE ASSESSMENT (LCA) OF SOLAR WATER HEATERS—A CASE STUDY

Solar water heaters use solar energy as the fuel and hence are environmentally friendly, as they do not generate any toxic emissions during their operation. There are, however, environmental burdens associated with solar water heaters due to the materials used and the fabrication that is involved. In order to improve the thermal performance of a built-in-storage system, Muneer and Asif developed a finned device employing stainless steel, with a collector area of 1 m^2 and capacity of 80 l. They conducted rigorous testing of the heater for a complete year to determine its performance under varying weather conditions in Pakistan. They also conducted LCAs to investigate its energy and environmental performance. Their findings are presented in Table 1.

Table 2 Installed solar thermal collector area for European countries (Muneer 2004)

Country	1	2	3	4	5
Germany	420	615	900	3.71	45.1
Greece	195	181	160	2.98	283.4
Austria	141	168	169	2.34	288.9
France	16	23	38	0.55	9.3
Spain	33	40	50	0.45	11.4
Denmark	22	27	40	0.32	60.6
Italy	16	18	17	0.31	5.5
Switzerland	30	26	27	0.26	36.1
Portugal	5	6	8	0.25	25
Netherlands	30	32	35	0.21	13.5
Sweden	9	18	13	0.21	23.9
United Kingdom	9	10	11	0.21	3.5
Finland	9	10	10	0.03	5.9
Belgium	2	2	3	0.02	2.4
Ireland	1	0.3	0.3	0.003	0.9
Total	914	1171	1488	11.9	

Column labeled 1: thousands of m^2 installed in year 1999. Column labeled 2: thousands of m^2 installed in year 2000. Column labeled 3: thousands of m^2 installed in year 2001. Column labeled 4: millions of m^2 installed, cumulative. Column labeled 5: cumulative installed m^2/thousand population.

The annual average daily incident solar irradiation on a 1 m^2 collector area of the heater, under the test conditions for the above site was 4.8 kWh. The average daily energy yield of the heater was found to be 2.65 kWh, thus demonstrating an efficiency of 55%.[12] It was found that over a service life of 20 years, the heater would provide 19.4 MWh of energy. The heater consequently has a potential of saving 1393 kg of CO_2 emission.

As part of this work, a life cycle cost assessment was also carried out. Compared to furnace oil systems, typically used within industrial settings, the solar water heater was found to have a payback period of 6 years.[8] Further work in this regard showed that the performance of the system could be significantly improved by employing aluminum collectors. Owing to the fact that aluminum is cheaper and has better thermal properties than stainless steel, the payback period of this type of system was found to be just over 3 years.

USER SURVEY AND MARKET GROWTH

The Energy Technology Support Unit (ETSU) undertook a survey of domestic users of solar water heaters. Users from all regions of the United Kingdom were surveyed on a number of issues ranging from the types of systems in use to economic viability and technical reliability of their systems. The majority of respondents (70%) registered their response as "very satisfied." Any initial dissatisfaction was commonly reported as being due to installation problems. It was also observed that post-1986 improvement in manufacturing technology had resulted in increased user satisfaction. On an average, respondents had paid £2500 for their system. The overall conclusion was that the systems had shown a high reliability, with 85% of the systems sold since 1974 still in use.[1]

Since the early 1970s, the efficiency and reliability of solar water heating systems have increased significantly while the cost has dropped. Improvements to materials, a rating system for consumers, and more attractive designs, have all helped to make systems more successful. Particularly, over the last ten years there has been a significant increase in the use of domestic solar water heaters across the world. In 2003, solar water heaters received 21% share of the total investment, U.S. $22 billion, in the renewable energy sector worldwide.[13] In 2004, China's solar water heating market grew by 30% as it installed well over 10 million m^2 of new solar thermal collector area—equivalent to 7 GWth. In terms of installed capacity per capita, Israel, Greece, and Austria are the world leaders. The solar water heating market in France grew by over 30% per year during the last 3 years. In Spain, the solar water heating market has grown at a rate of 15% over the last 6 years while an annual growth target of 65% has been set until the year 2010. Similarly, other European countries are also experiencing an upward trend as shown in Table 2.[14] Reported figures indicate a healthy growth in 2004 (compared with 2003) in several countries, even though the total market is still small. Belgium grew by 62%, Estonia by 67%, Hungary by 50%, Ireland by 67%, Malta by 41%, Portugal by 67%, and Slovenia by 64%. In the United States, solar water heating is becoming increasingly popular in domestic and swimming pool applications.[15] From 1996 to 2004, for example, the Hawaiian Electric Company (HECO) has seen more than 25,000 solar water heater (SWH) systems installed within its customer base.[16]

CONCLUSIONS

Solar water heating is an energy efficient, cost effective and environmentally friendly renewable energy technology that has a huge potential in domestic and industrial applications. For most countries around the world, solar water heaters can provide up to two-thirds of the total hot water requirements while cutting down the energy cost and the associated environmental impacts. Due to its favorable characteristics, solar water heating is becoming increasingly popular across the world. In 2003 approximately 4.6 billion USD were invested in solar water heating, accounting for 21% of the total investment made within the renewable energy sector. China, a leading player in the solar water heating market, saw its solar water heating market grow by 30% in the year 2004. On the other hand, several European countries are experiencing an even healthier growth. On the basis of a case study presented in this entry, it may also be noted that the payback periods for built-in-storage heaters using stainless steel and aluminum may be expected to be, respectively, around 6 and 3 years.

REFERENCES

1. Muneer, T. Solar Eenergy In *KEMPS Engineering Year Book*; Miller Freeman: Kent, England, 2000.
2. http://www.toolbase.org/tertiaryT.asp?DocumentID=3216&CategoryID=949 (accessed October 2005).
3. Vijayaraghavan, S. In *Encyclopaedia of Energy*, Vol. 5; Academic Press: London, 2004.
4. Muneer, T. Effect of design parameters on performance of built—in storage solar water heaters. Energy Conversion Manag. **1984**, *3*, 277–281.
5. Kreider, J.F.; Kreith, F. Non-Concentrating Solar Thermal Collectors Solar Energy Handbook. In *Solar Energy Handbook*; McGraw-Hill: New York, 1981.
6. Liu, W.; Davidson, J.H.; Kulacki, F.A. Thermal characterisation of prototypical integral collector storage system with immersed heat exchangers. J. Solar Energy **2005**, *127*, 21–28.

7. Cruza, M.S.; Hammond, G.P.; Reisa, A.S. Thermal performance of a trapezoidal-shaped solar collector/energy store. Appl. Energy **2002**, *73*, 195–212.
8. Munner, T.; Maubleu, S.; Asif, M. Prospects of solar water heating for textile industry in Pakistan. Renew. Sust. Energy Rev. **2006**, *10*, 1–23.
9. *Product Range*, Viessmann Werke GmbH & Co KG, Germany, 1997
10. Duffie, J.; Beckman, W. *Solar Engineering of Thermal Processes*, 2nd Ed.; Wiley: New York, 1991.
11. *Heat your Water With the Sun: A Consumer Guide*; U. S. Department of Energy; Washington, DC, U.S.A.; 2003.
12. Asif, M.; Muneer, T. Life cycle assessment of solar water heating in Pakistan; *Building Services Engineering Research and Technology*; in press.
13. Annual investment in renewable energy: 1995–2003; *Renewable Energy World*; February 2005.
14. Muneer, T. *Solar Radiation and Daylight*, 2nd Ed.; Elsevier Butterworth Heinemann: Oxford, U.K.; 2004.
15. Solar Thermal Systems on the Up; *Renewable Energy World*; July 2005.
16. http://go.ucsusa.org/clean_energy/renewable_energy/page.cfm?pageID=1595 (accessed October 2005).

Res–Solar

Solid Waste to Energy by Advanced Thermal Technologies (SWEATT)

Alex E. S. Green
Green Liquids and Gas Technologies, Gainesville, Florida, U.S.A.

Abstract

Solid waste (SW), now mostly wasted biomass, could fuel approximately ten times more of the United States's increasing energy needs than it currently does. At the same time, it would create good nonexportable jobs and local industries. Twenty-four examples of wasted or underutilized solids that contain appreciable organic matter are listed. Estimates of their sustainable tonnage exceed two billion dry tons. Now usually disposal problems, most of these SWs can be pyrolyzed into substitutes for or supplements to expensive natural gas (NG). The large proportion of carbon-dioxide-neutral plant matter in the list would reduce greenhouse problems. Pyrolysis—heating to high temperatures without oxygen—converts such SW into a medium-heating-value gaseous fuel, usually with small energy expenditure. With advanced gas cleaning technologies the pyrogas can be used in high-efficiency gas turbines or fuel-cell systems. This approach has important environmental and efficiency advantages with respect to direct combustion in boilers and even air-blown or oxygen-blown partial combustion gasifiers. Because pyrolysis is still not a predictive science, the Clean Combustion Technology Laboratory (CCTL) has used an analytical semiempirical model (ASEM) to organize experimental measurements of various product $\{C_aH_bO_c\}$ yields vs temperature (T) for dry ash, nitrogen, and sulfur free (DANSF) feedstock having various weight percentages (wt%) of oxygen [O], and hydrogen [H]. With this ASEM, each product is assigned five parameters (W, T_0, D, p, q) in a robust analytical $Y(T)$ expression to represent yields vs temperature of any specific product from any specified feedstock. Patterns in the dependence of these parameters upon [O], [H], a, b, and c suggest that there is some order in pyrolysis yields that might be useful in waste-to-energy conversion (WEC) systems to optimize their throughput. An analytical cost estimation (ACE) model is used to calculate the cost of electricity (COE) vs the cost of fuel (COF) for a SW integrated gasifier combined cycle (IGCC) system for comparison with the COE vs COF for a natural gas combined cycle (NGCC) system. It shows that at high NG prices, SW can be changed from a disposal-cost item to a valuable asset. Comparing COEs when using other SW-capable technologies are also facilitated by the ACE method. Implications of this work for programs that combine conservation with waste-to-energy conversion in efforts to reach Zero Waste are discussed.

SOLID FUELS AND SOLID WASTE

In 1940, when Britain was fighting a ruthless and apparently unstoppable Hitler, Winston Churchill offered only "blood, sweat, toil, and tears" to unite Britain's political factions. At this time in our history we are excessively (60%) reliant on foreign sources for our liquid fuels and are increasingly importing our gaseous fuels (now $>15\%$). Our country is shedding blood in its efforts to stabilize regions of the globe that supply these premium fuels. Yet the United States is well endowed with solid fuels in the form of coal and oil shale, and substantial quantities of renewable but wasted solids. In this paper, in continuation of a long search for alternatives to oil,[1–10] our focus is on converting our solid waste to energy by advanced thermal technologies (SWEATT).

Keywords: Solid waste; Pyrolysis; Gasification; Cost of electricity; Pyrolysis products.

Table 1 is a list of the United States's abundant supply of wasted solids or SW whose organic matter can be made into gaseous and liquid fuels. With recent high natural gas (NG) prices and for technical reasons that will become obvious, this entry will concentrate on advanced thermal technologies (ATT) for the conversions of SW to gaseous fuels. Advanced thermal technologies conversions of coal to liquid and gaseous fuels involve similar technical considerations, but coal to liquid or gas technologies have the attention of many government, business, and engineering personnel. Solid waste to energy by advanced thermal technologies has the attention of only a few in the United States.

In the United States, most of the categories in Table 1 would now be called "biomass" in part because "solid waste" has a bad public image, bringing to mind old incinerators belching black smoke. However, with advances in thermal technologies and gas cleanup systems now being successfully applied in Japan and the European Union (EU),[11] SWEATT deserves a new image. It not only addresses the United States's very urgent need for

Table 1 Wasted solids that could be used as a component of the united state's primary energy supply

Waste type	Million dry tons
Agricultural residues	~0.98
Forest under-story and forestry residues	~0.40
Hurricane debris	~0.04
Construction and deconstruction debris	~0.02
Refuse derived fuels	~0.10
Urban yard waste	~0.02
Food serving and food processing waste	~0.07
Used newspaper and paper towels	~0.02
Used tires	~0.05
Energy crops on under-utilized lands	~0.05
Ethanol production waste	~0.02
Anaerobic digestion waste	~0.01
Bio-oil production waste	~0.01
Waste plastics	~0.03
Infested trees (beetles, canker, spores)	~0.02
Invasive species (cogon-grass, melaluca)	~0.02
Plastics mined when restoring landfills	~0.03
*Bio-solids (dried pelletized sewage sludge)	~0.04
*Poultry and pig farm waste	~0.02
*Water plant-remediators (algae, hydrilla.)	~0.01
*Muck pumped to shore to remediate lakes	~0.01
Manure from cattle feed lots	~0.01
Plants for phyto-remediation of toxic sites	~0.01
Treated wood past its useful life	~0.01
Total	~ 2 billion dry tons

Table 1 is a list of potential local sources of useful nonconventional fuels cited in the author's conference presentations with recent emphasis on sources available in Florida.Items marked with * help in water remediation and the ~ denotes estimated values.

alternative fuels, but also could mitigate air and water pollution problems. The large carbon dioxide-neutral plant matter components in Table 1 can help in greenhouse mitigation. The great diversity of physical and chemical characteristics in Table 1 implies that the world needs omnivorous feedstock converters (OFCs) to change these solid fuels into much more usable liquid or gaseous fuels.

Fig. 1 is a conceptual illustration of an OFC adapted from several previous CCTL papers[8–10]in which a SW pyrolyzer–gasifier–liquifier is co-utilized with a natural gas combined cycle (NGCC) system, as discussed in "ASEM and Pyrolysis."

PROPERTIES OF SOLID FUELS AND SOLID WASTE

The declining resources of domestic liquid and gaseous fuels are the greatest energy problems facing the United States and many other countries today. Yet, the United States is particularly well endowed with solid fuels in the form of various coals, peat, biomass, and SW. Table 2 shows major ranks of coals as well as of peat, wood and cellulose, and their ultimate and proximate analyses as measured by industry for over a century. The numbers listed in columns 2, 3, and 4 essentially apply to ideal carbon, hydrogen, and oxygen (CHO) materials by correcting measurements to be dry, ash, sulfur, and nitrogen free (DASNF). For such material [C] = 100–[O]–[H] so [C] becomes a variable dependent upon the values of [H] and [O]. Column 5 gives the higher heating value (HHV) in MJ/kg as measured with standard bomb calorimeters after allowing for the minor components.

Fig. 2a is mainly a plot of [H], the wt% of hydrogen (solid diamonds with values read on the left scale) vs [O], the wt% of oxygen, for 185 representative DANSF CHO materials taken from ultimate analysis data available in the technical literature. The bottom scales give conventional coal ranks, some potential names for the biomass region, and some names that might foster more friendly discussions between the coal and biomass communities. This [H] vs [O] coalification plot shows that apart from the anthracite region, all natural DANSF feedstock have [H] values that are close to 6%. The near constancy of [H] together with the relationship [C] = 100 − [O]–[H] for DASNF feedstock imply a linear decline of [C] with increasing [O]. The smooth trend of properties with [O] provide strong reasons for regarding peat and biomass simply as lower-rank coals or the various coal ranks simply as aged forms of biomass. The diagram suggests that the natural solid fuels could be ranked simply by [O] to replace the different ranking systems of various countries (a Tower of Babel!). Using 34-O for peat—called "turf" in Ireland—might help temper the "turf wars" in fuel sector competitions and in energy/environmental confrontations on the use of our available solid fuels.

Higher heating values of various fuels measured with calorimeters usually are reported with proximate analyses. Approximate HHVs in MJ/Kg for the seven representative CHOs are given in Column 5 in Table 2. Column 6 gives

Solid–Sus

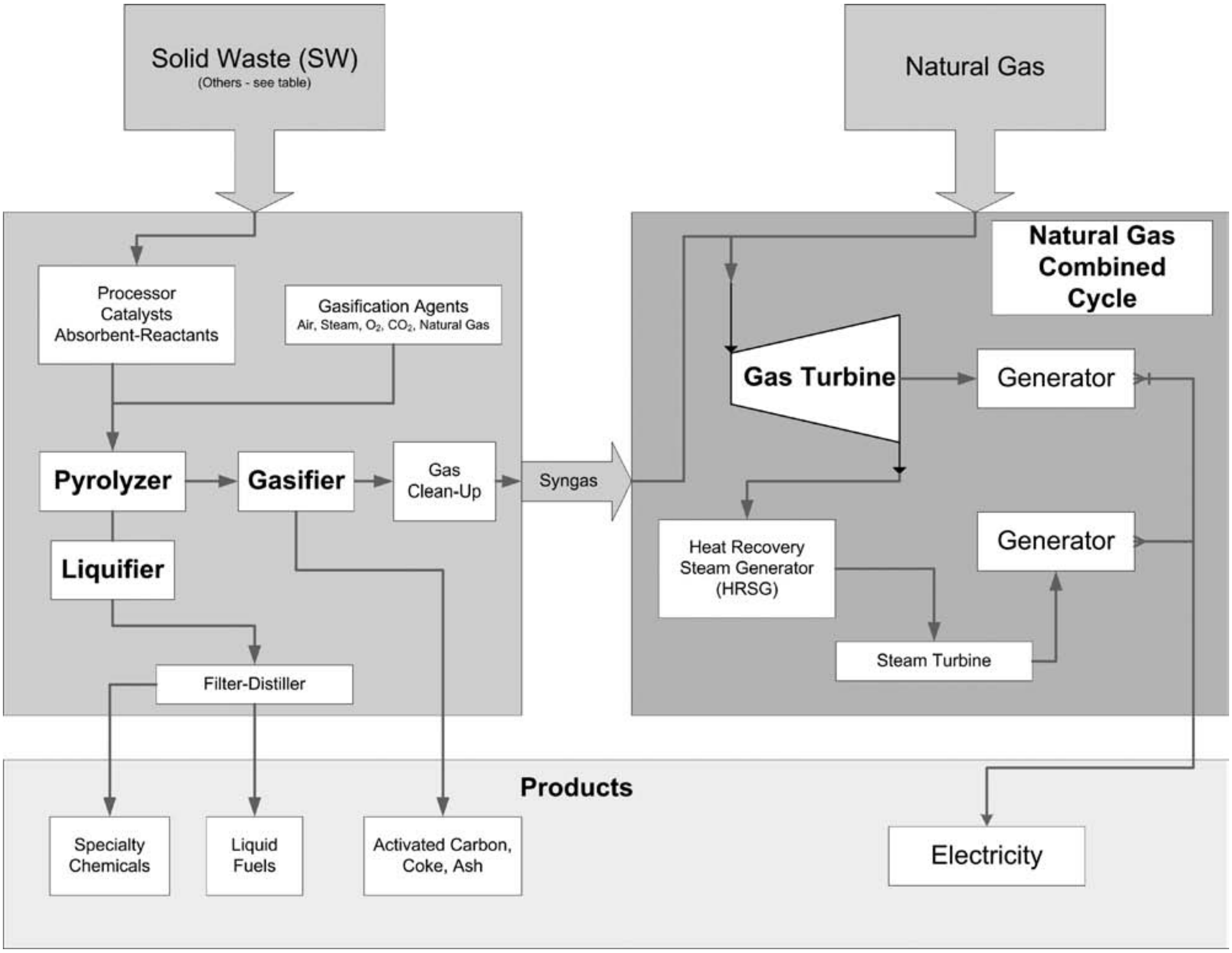


Fig. 1 Diagram of the omnivorous feedstock converter (OFC) illustrating the addition of a solid waste (SW) system to an existing natural gas combined cycle (NGCC) plant to create an effective SWCC system.
Source: Adapted from Ref. [10].

representative total volatiles, V_T, as determined by an American Standard Test Measurement Method (ASTM). A solid sample is heated (pyrolyzed) in an inert atmosphere using a platinum crucible at 950°C for 7 min. The wt% loss between the sample and its char is the total volatile yield. Then the balance from 100% represents the weight of the fixed carbon (FC) plus ash. When this residual is burned, the remainder is the ash wt%. An empirical analytical formula is given in the caption to represent general trends of total volatiles along nature's coalification curve. Note the rapidly increasing trend in V_T from low [O] materials to high [O] materials. The numbers in Column 7 of Table 2 represent $FC = 100 - V_T$, the fixed carbon for pure CHO materials after all volatiles are driven off.

Columns 8 and 9 of Table 2 give the relative physical density and relative energy density of the various natural solid fuels, which are important factors in handling and transportation costs. Columns 10 and 11 relate to the reactivity and H and OH free radical generated in the combustion of these various solid fuels, which have strong influence on rates of reaction. The last column gives the proposed quantitative [O] ranking system for solid fuels that lie along nature's coalification curve.

The HHV of a solid fuel is, perhaps, the most important variable in solid fuel use and solid fuel conversion to liquid and gaseous fuels. A simple form of Dulong's formula, suitable for mental calculations, that captures the main trends is

$$\begin{aligned} \text{HHV} &= [\text{C}]/3 \\ &\quad + 1.2[\text{H}]–[\text{O}]/10 \text{ in MJ/Kg} \\ &\quad \text{where M} = 1{,}000{,}000 \end{aligned} \tag{1}$$

Divide this number by 2.3 to get the corresponding HHV in MBtu, where $M = 1000$. Note that the carbon

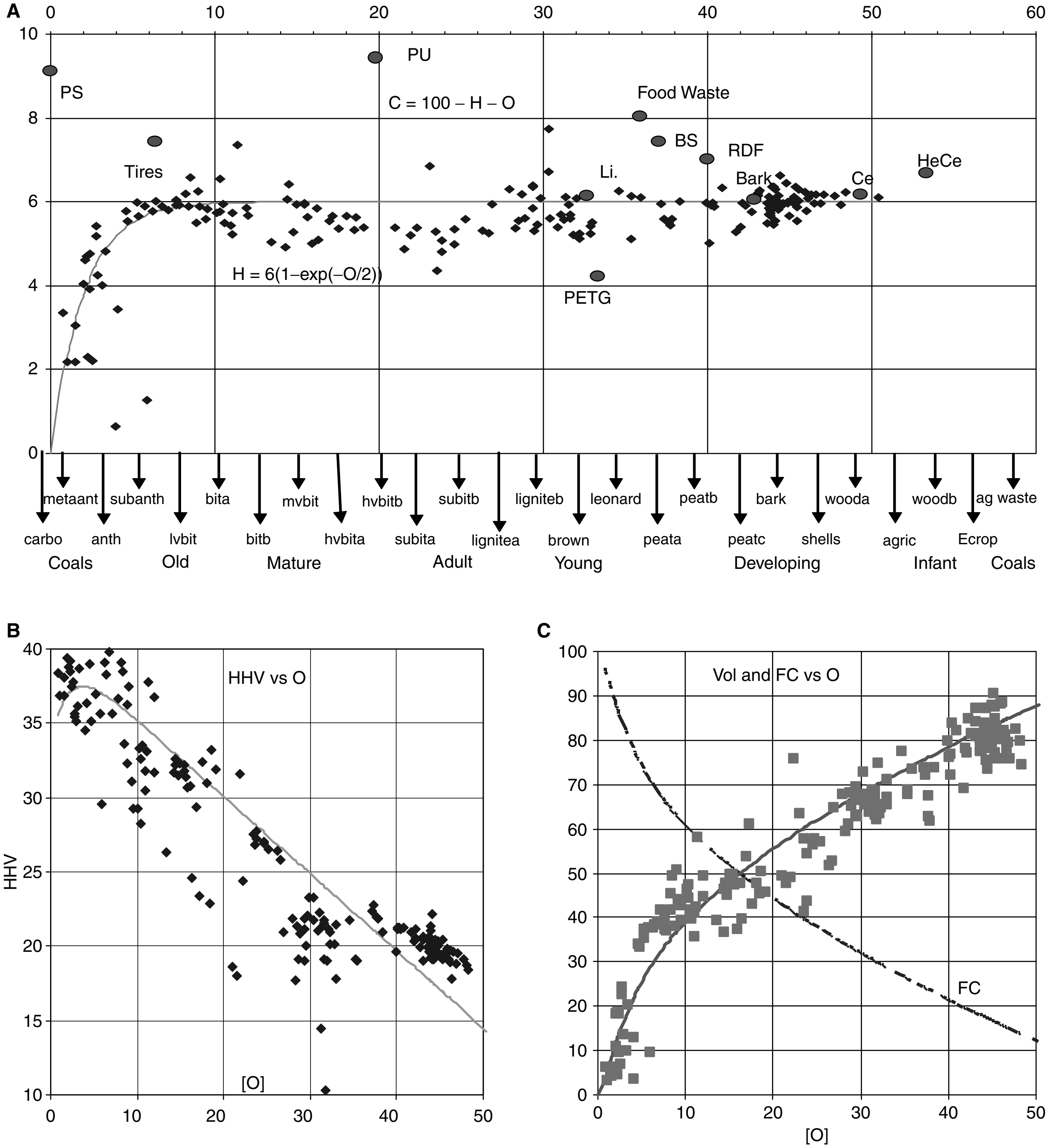


Fig. 2 (A) Weight percentages of hydrogen [H] vs [O] for 185 DANSF carbonaceous materials (black diamonds) vs oxygen wt%. Classification labels are given at the bottom scale and [O] values on top scale. Adapted from Ref. [4]. (B) Higher heating values (HHV) of 185 carbonaceous materials (corrected to DANSF) vs [O]. The smooth curve represents HHV = ([C]/3 + 1.2[H]–[O]/10). (C) Total volatile weight percentages vs [O] for 185 DASNF carbonaceous materials (squares) from proximate analysis. The curve through the data points satisfies $V_T = 62([H]/6)([O]/25)^{1/2}$. The analytic fixed carbon (FC) is shown.

Table 2 Properties of fuels along natures coalification path

	Ultimate analysis			Proximate analysis			Other properties				
Name	**C**	**H**	**O**	**HHV**	**VT**	**FCCh**	**Dens**	**E/vol**	**RelchR**	**H,OH Rad**	**O-Rank**
Anthracite	94	3	3	36	7	93	1.6	58	1.5	v. low	3-O
Bituminous	85	5	10	35	33	67	1.4	49	5	low	10-O
Sub bitum	75	5	20	30	51	49	1.2	36	16	med	20-O
Lignite	70	5	25	27	58	42	1	27	50	interm	25-O
Peat	60	6	34	23	69	31	0.8	18	150	high	34-O
Wood	49	7	44	18	81	19	0.6	11	500	v. high	44-O
Cellulose	44	6	50	10	88	12	0.4	9	1600	v v.high	50-O

energy term ([C]/3) usually is much larger than the hydrogen energy contribution (1.2[H]). Oxygen contributes negatively in part because the more [O] implies less [C] and in part because of the subtractive term −[O]/10. The larger points on Fig. 2a give the [H], [O] positions of lignin (6.1,32.6), cellulose (6.2,49.4), and hemi-cellulose (6.7,53.3), the three main components of all plant matter. Also shown in Fig. 2a are the [H] and [O] coordinates of several materials that are present in SW. These depart substantially above and below the coalification curve. Not shown are petroleum and polyethylene, which would lie at [14.2, 0].

Fig. 2b shows the pattern of HHVs vs [O] along the coalification path. The DuLong formula given in the caption is a simple compromise between those used in the coal and biomass sectors. Fig. 2c shows the total volatiles (V_T) for the CHO materials vs [O] mostly for materials close to the coalification path. These values are determined by standard proximate analysis procedures that measure the weight loss of a sample after exposure to 950°C for 7 min in an anoxic medium. It should be obvious that in the high [O] region pyrolysis is substantially equivalent to gasification. Detailed studies point to the fact that small departures of [H] from the coalification curve have large impacts on volatile content.

The three diagrams in Fig. 2 all indicate the importance of the [O] in determining the fuel properties of natural substances. CCTL studies indicate that the [H] dimension is also very important in determining volatile content and that small deviations of [H] from the smooth coalification path have a large impact on the volatile release. Table 3 gives a compact list of heating values of various wastes, fuels, and plastics in units of MBtu/lb.

GLOBAL AND U.S. PRIMARY ENERGY SUPPLIES

Fig. 3a presents an overview of the world total primary energy supply (see the International Energy Agency Web site) at the opening of the millennia. Among the major sources, combustible renewables and waste (CRW, mostly biomass) need only be doubled to be competitive with coal and NG, and tripled to be competitive with petroleum. Already, the category CRW is almost a factor of 2 greater than nuclear. On the other hand, wind and solar must grow by factors of more than 100 to become major global energy supplies. This global total primary energy supply (TPES) picture is not representative of the industrial world, particularly the United States today. Fig. 3b shows the subdivisions of the U.S. TPES in 2005, in quadrillion Btu or quads (see the U.S. Energy Information Agency Web site).[12] Because total consumption is now very close to 100 quads, the numbers might also be considered to be

Table 3 Heating values of MSW components, fuels, and plastics in 1000 Btu/lb

Component	As recd	Dry	Component	Dry
Wastes			Fuel	
Paper and paper products			*Hydrocarbons*	
Paper, mixed	6.80	7.57	Hydrogen	60.99
Newsprint	7.97	8.48	Natural gas	20.00
Brown paper	7.26	7.71	Methane	23.90
Trade magazines	5.25	5.48	Propane	21.52
Corrugated boxes	7.04	7.43	Ethane	22.28
Plastic-coated paper	7.34	7.70	Butane	21.44
Waxed milk cartons	11.33	11.73	Ethylene	21.65
Paper food cartons	7.26	7.73	Acetylene	21.50
Junk mail	6.09	6.38	Naphthalene	17.30
			Benzene	18.21
Domestic wastes			Toluene	18.44
Upholstery	6.96	7.48	Xylene	18.65
Tires	13.80	13.91	Naptha	15.00
Leather	7.96	8.85	Turpentine	17.00
Leather shoe	7.24	7.83		
Shoe, heel, and sole	10.90	11.03	*Oils*	
Rubber	11.20	11.33	No. 1 (Kerosene)	19.94
Mixed plastics	14.10	14.37	No. 2 (Distillate)	19.57
Plastic film	—	13.85	No. 4 (VL residual)	18.90
Linoleum	8.15	8.31	No. 5 (L residual)	18.65
Rags	6.90	7.65	No. 6 (residual)	18.27
Textiles	—	8.04		
Oils, paints	13.40	13.40	*Alcohols*	
Vacuum-cleaner dirt	6.39	6.76	Methanol	10.26
Household dirt	3.67	3.79	Ethanol	13.15
Food and food waste			Plastics	
Vegetable food waste	1.79	8.27		
Citrus rinds and seeds	1.71	8.02	Polyethylene	19.73
Meat scraps (cooked)	7.62	12.44	Polystyrene	16.45
Fried fates	16.47	16.47	Polyurethane	11.22
Mixed garbage 1	2.37	8.48	Polyvinyl chloride	9.78
			PVC (pure resin)	7.20
Coals			Polyvinylidene chloride	4.32
Low-vol. bituminous	—	15.55	Polycarbonate	13.31
Med.-vol. bituminous	—	15.35	Cellulose	7.52
High-vol. bituminous	12.25	14.40	Polypropylene	20.22
Subbituminous	9.90	12.60	Polyester	12.81
Lignite	7.30	11.45		
Anthracite	—	14.00		

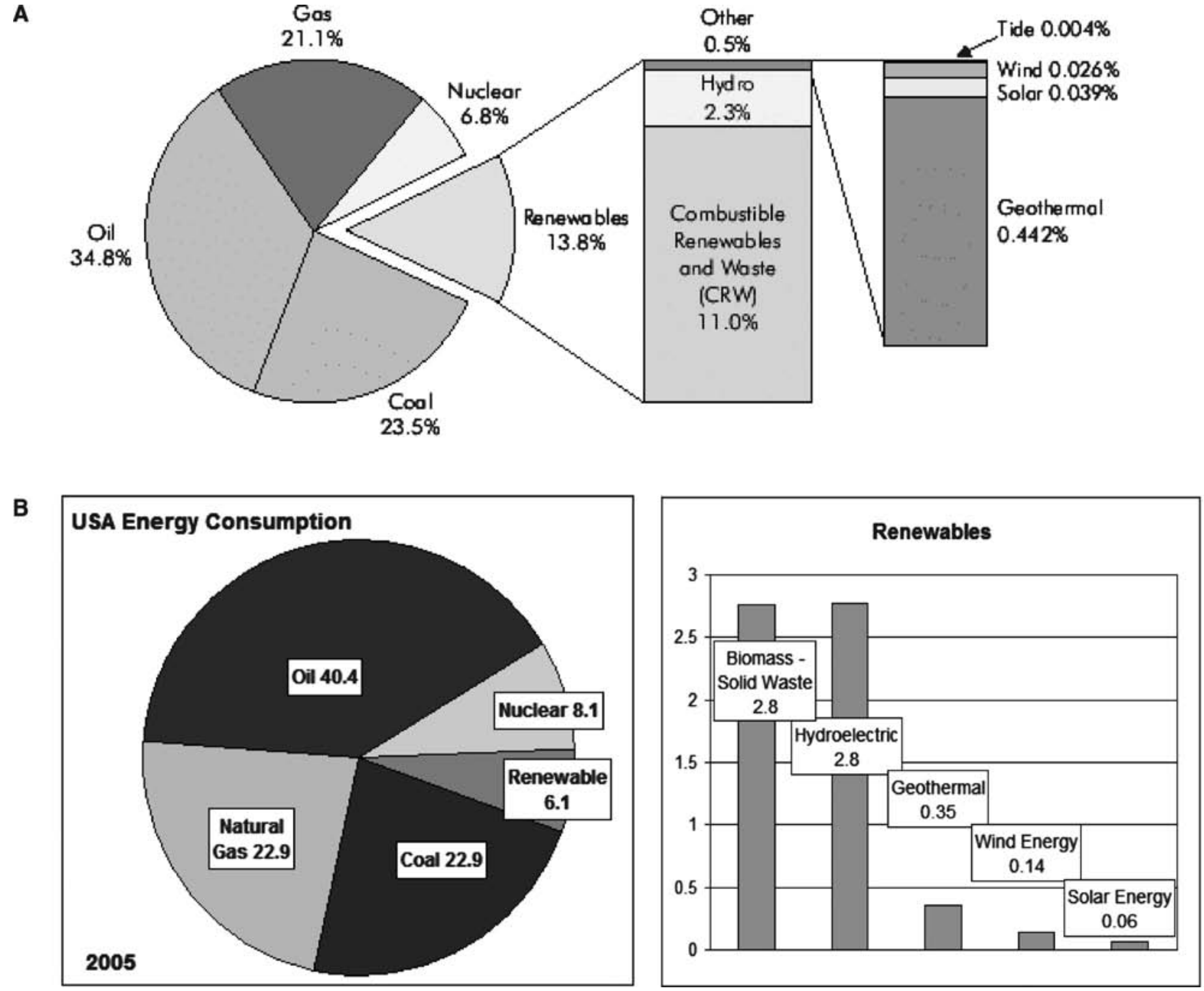


Fig. 3 (A) Total primary energy supply for the globe at the millennium (IEA Web site), (B) (Left) total 2005 annual U.S.A. energy consumption of primary energy sources in quads (Right) renewables.

approximate percentages of U.S. energy consumption. It is seen that more than 40% of our energy consumption is in the form of petroleum, consumed mainly in our transportation sector. Without doubt the biggest energy problem faced by the United States today is the need to find alternatives to oil.[1–3] In the 1970s and early 1980s, the United States focused heavily on alternatives to oil in the utility sector. The alternatives first were pulverized coal plants and, in the late 1980s and 1990s, NGCC systems. At this time, the United States' focus should in part be on the developing alternatives to NG for electricity generation via the use of ATT. It should be noted, however, that ATT can also make significant contributions to the solution of our liquid fuel problem in the transportation sector.[3]

It is important to differentiate secondary energy supplies (SES) from the primary energy supplies (PES) shown in Fig. 3. Secondary energies supplies include steam, syngas, reactive chemicals, hydrogen, charges in batteries, fuel cells, and other energy sources that draw their energy from PES. If an SES is converted to another type of energy—say, mechanical energy—via a steam turbine, the mechanical energy becomes a tertiary energy supply (TES). This TES can be converted to electrical energy via magnetic generators, in which case the electricity is a quaternary supply (QES). In the case of electricity, the many conversions are usually justified because electricity can readily be distributed by wire and has so many uses as a source of energy for highly efficient electric motors, illumination systems, home appliances, computers, etc.

A debate is under way in many communities as to whether increasing electricity needs should be met with solid fuels—particularly coal—via conventional steam and steam turbine generator systems or via conversion to a gaseous fuel to fuel integrated gasifier combined cycle (IGCC) systems. Granting that the steam turbine route has had many advances over the past century, our thesis is that converting the solid fuel to gaseous fuel is the ATT route of the future. The ATT route is driven not only by environmentally acceptable waste disposal needs and increased needs for electricity, but also by the need for liquid and gaseous fuels. A number of petroleum resource experts recently advanced the date when the globe's supply of oil and NG will run out. The prices of oil and NG now reflecting this drawdown are already high enough that conversion of organic matter in SW to liquid and gaseous fuels makes economic sense. We should recognize, however, that for the most part, cartels—not free markets—govern fuel prices. Thus, we should not abandon alternative fuels efforts whenever cartels, for their interests, lower prices.

The SWs listed in Table 1, mostly consisting of biomass, now constitute a minor component ($\sim$2.8%) of the United States' annual TPES. This wasted material, however, could in the near term become a major ($>$25%) component comparable to coal and NG, both now at about 23%. Essentially, the United States now consumes about 100 quadrillion Btu, only about 2.8% of which currently come from SW. The other renewables—hydroelectric (2.8%), geothermal (0.35%), wind (0.14%), and solar (0.06%)—have much further to go than SW before becoming a major primary energy source in the United States. Because SWEATT is based on locally available SW, it would also create good nonexportable local industries and jobs while mitigating serious U.S. energy import and waste disposal problems.

An Oak Ridge National Laboratory study[13] estimates the sustainable supply of the first few biomass categories in Table 1 at about 1.4 billion dry tons. The remaining categories should readily bring the total sustainable U.S. SW available to more than two billion dry tons. Assuming a conservative HHV of 7500 Btu/lb, a simple calculation shows that with SWEATT technologies similar to those that are now in place in Japan, U.S. SW contribution to its primary energy supply could reach the 25% level.

ADVANCED THERMAL TECHNOLOGIES

The largest SW-to-energy systems in operation today are direct combustion municipal solid waste (MSW) incinerators[14] with capacities in the range of 1000–3000 tons per day. In such mass burn systems, the organic constituents of the SW are combusted (in a sense, converted) into the gaseous products CO_2 and H_2O. These have no fuel value but can be carriers of the heat of combustion, as in coal and biomass boiler-furnace systems. Along with the flame radiation, these gases transfer heat to pressurized water to produce pressurized steam that drives a steam turbine-driven electric generator. The steam also can serve as a valuable secondary energy supply (SES) to distribute heat for heating buildings, industrial processes, etc. The production and use of steam, along with the steam engine, launched the Industrial Age, and various steam-driven systems have reached a very high level of refinement, including in waste-to-energy systems.[14]

Solid waste to energy by advanced thermal technologies systems do not involve direct combustion and the use of the heat released to raise steam; rather, the SW is first converted to a gaseous or liquid fuel. Then this fuel serves as an SES that can be combusted in efficient internal combustion engines, combustion turbines, or (in the future) in fuel cells, none of which can directly use solid fuels. Over the past century, automotive and aircraft developments have pushed internal combustion engines (ICE) and gas turbines (GT) to very high levels of efficiency. Furthermore, with the use of modern high-temperature GTs in NG-fired combined cycle (NGCC) systems, the heat of the exhaust gases can be used with a heat recovery steam generator (HRSG) to drive a steam turbine. Alternatively, the HRSG can provide steam for combined heat and power (CHP) system that can effectively make even greater use of the original solid fuel energy.

If one considers the United States' heavy dependence on foreign sources of liquid and gaseous fuels, the most challenging technical problem facing us today should be recognized as the development and implementation of efficient ways of converting our abundant domestic solid fuels to more useful liquid and gaseous fuels. In view of the diversity of feedstock represented in municipal or institutional SW, any successes in SWEATT would advance this more general quest. In effect, the United States and the world need an OFC such as is illustrated in Fig. 1. Here, the right block represents a typical gas-fired combined cycle system, whereas the left block represents a conceptual omnivorous conversion system that can convert any organic material into a gaseous or liquid fuel.

GROSS COMPARISONS OF ATT OUTPUTS

First, we will consider the gross nature of the output gas from biomass and cellulosic-type material, the major organic components of most solid-waste streams. Apart from minor constituents such as sulfur and nitrogen, the cellulosic feed types are complex combinations of carbon, hydrogen, and oxygen in combinations such as ($C_6H_{10}O_5$) that might serve as the representative cellulosic monomer.

Advanced thermal technologies systems may be divided into (1) air-blown partial combustion (ABPC) gasifiers, (2) oxygen-blown partial combustion (OBPC) gasifiers, and (3) pyrolysis (PYRO) systems. The three approaches for converting waste to a gaseous fuel have

Solid–Sus

many technical forms, depending on the detailed arrangements for applying heat to the incoming feed and the source of heat used to change the solid into a gas or liquid.

Let use "producer gas" as a generic name for gases developed by partial combustion of the feedstock with air, as in many traditional ABPC gasifiers that go back to Clayton's coal gasifier of 1694. We will use "syngas" for gases developed by partial combustion of the feedstock with oxygen, as in OBPC gasifiers, which are mainly a development of the 20th century. We will use "pyrogas" for gases developed by anaerobic heating of the feedstock, such as in indirectly heated (PYRO) gasifiers. Our objective is to replace NG that has HHV ~ 1000 Btu/ft^3 = 1 MBtu/ft^3 (here, $M = 1000$).

When an ABPC gasifier is used with cellulosic materials (cardboard, paper, wood chips, bagasse, etc.), the HHV of biomass producer gas is very low (100– 200 Btu/ft^3) for two reasons: (1) The main products are CO that has a HHV of 322 Btu/ft^3 and CO_2 and H_2O that have zero heating values and (2) the air nitrogen substantially dilutes the output gas.

The syngas obtained from biomass with an OBPC gasifiers is better— ~ 320 Btu/ft^3 because it is not diluted by the atmospheric nitrogen. It is still somewhat lower than the feedstock molecules, however, because of the partial combustion. The oxygen separator is a major capital-cost component of an OBPC gasifier.

With a PYRO system, the original cellulosic polymer is first broken into its monomers, leading to some CO, CO_2, and H_2O along with paraffins (CH_4, C_2H_6, and C_3H_8…), olefins (C_2H_4, C_3H_6,…) and oxygenated hydrocarbons: carbonyls, alcohols, ethers, aldehydes and phenols, and other oxygenated gaseous products. Cellulosic pyrogas can have heating values in the 400 Btu/ft^3 range.

Hydrocarbon plastics such as polyethylene and polyolefins in general are among the most predominant plastics in many SW streams. Thus, one might use (C_2H_4) as representative of the monomers in the plastic component of MSW or refuse-derived fuels (RDF). Polyethylene pyrolysis products include H_2, olefins, paraffins, acetylenes, aromatics (Ar), and polynuclear aromatics (PNA). On a per-unit weight basis, all but H_2 have gross heating values in the range 18–23 MBtu/lb ($M = 1000$), similar to oil, whereas H_2 has a gross heating value of 61 MBtu/lb. On a per-unit volume, basis all polyethylene pyrolysis products have gross heating value ranging from 1 to 5 MBtu/ft^3, whereas H_2 is 0.325 MBtu/ft^3 = 325 Btu/ft^3. Natural gas typically is about 1 MBtu/ft^3. Thus, we would expect the pyrogas from polyethylene to have a gross heating value comparable to or greater than that of NG, and much greater than that of cellusosic pyrogas.

In summary, because cellulosic feedstock is already oxygenated as compared with pure hydrocarbon plastics, its pyrogas, syngas, and producer gas will all have considerably lower heating values than the corresponding gases from hydrocarbon feedstock. From the viewpoint of maximizing the HHV of SW-derived gas, PYRO gasification scores better than OBPC gasification, which scores better than ABPC gasification.

ASEM AND PYROLYSIS

Proximate analyses of coal and biomass measured for more than a century provide extensive data on total volatile content. A predictive method for identifying the molecules in these volatiles is still not available, however, despite the fact that such knowledge could provide a fundamental understanding of humankind's oldest technology: the use of fire. For control and application of a pyrolysis system, it would be useful to have at least an engineering-type knowledge of the expected yields of the main products from various feedstock subjected to anaerobic thermal treatment.

In most attempts to find the systematic of pyrolysis yields of organic materials such as coal and biomass, including the initial CCTL studies,[15–20] it has been customary to characterized the feedstock by its atomic ratios $y = \mathrm{H/C}$ and $x = \mathrm{O/C}$. In its recent studies,[21–27] the CCTL has found it more advantageous to work with the weight percentages [C], [H], and [O] of the feedstock after correcting to dry, ash, sulfur, and nitrogen free (DASNF) conditions (i.e., pure CHO materials). These were attempts to find some underlying order of pyrolysis yields of any product $C_aH_bO_c$ vs the [O] and [H] of the DASNF feedstock and the temperature (T) and time (t) of exposure. In organizing CCTL pyrolysis data as well as data in the literature, the CCTL has developed an analytical semi-empirical model (ASEM) that has been useful for several applications of pyrolysis.[19–27] Some progress has been made in including the time dimension, but much more work remains. When the time dimension is not a factor, the yields of each product for slow pyrolysis (or fast pyrolysis at a fixed time) are represented by

$$Y(T) = W[L(T : T_o, D)]^p[F(T : T_o, D)]^q \quad (2)$$

where

$$L(T : T_o, D) = 1/[1 + \exp((T_o - T)/D)] \quad (3)$$

and

$$F(T : T_o, D) = 1 - L(T)$$

$$= 1/[1 + \exp((T - T_o)/D)] \quad (4)$$

Here, $L(T)$ is the well-known logistic function that is often called the learning curve. Thus, its complement $F(T) = 1 - L(T)$ might be called the forgetting curve. For engineering applications, this curve-fitting approach provide a more robust and convenient means for organizing pyrolysis data than traditional methods that use

Solid–Sus

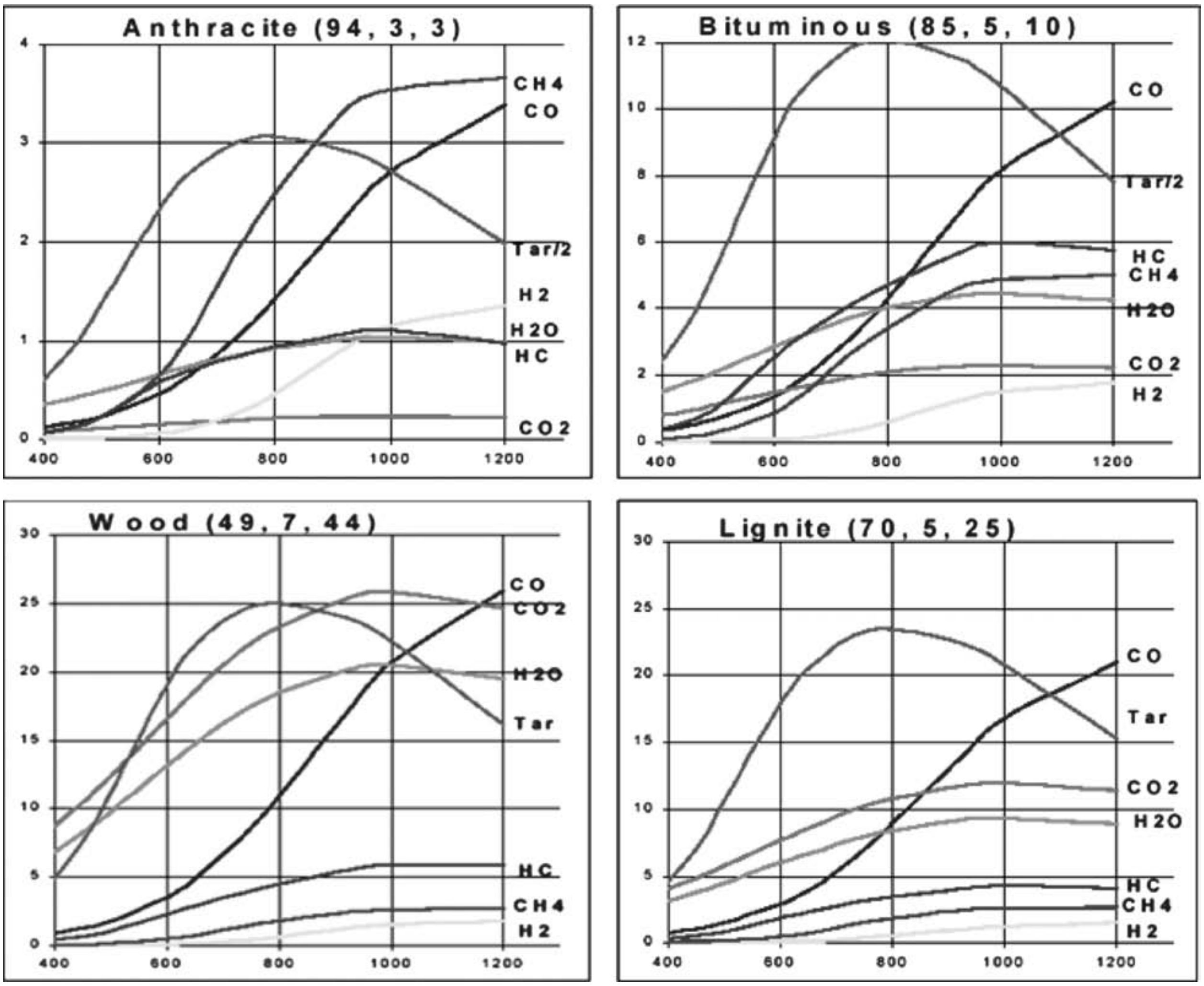


Fig. 4 Wt% yields vs temperature (in °C) from pyrolysis of anthracite, bituminous, lignite, and wood with ([C], [H], [O]) as shown. HC represents C2 and C3 gasses, BTX, phenol and cresol.
Source: Adapted from Ref. [10].

conventional Arrhenius reaction rate formulas.[28] In the ASEM, each product is assigned five parameters (W, T_0, D_0, p, q) to represent its yield-vs-temperature profile. The objective has been to find how these parameters depend on the [H] and [O] of the feedstock and the a, b, c of the $C_aH_bO_c$ product for the data from particular types of pyrolyzers. Studies by Xu and Tomita (XT)[29] that gave data on 15 products from 17 coals at six temperatures have been particularly helpful in revealing trends of the parameters with [O] and [H]. In applying the ASEM to the CCTL data collection, the XT collection, and several other collections, a reasonable working formula was found for the yield of any abc product for any [O], [H] feedstock. It was given by

$$Y(C_aH_bO_c) = W_{abc} z^{\alpha} h^{\beta} x^{\gamma} [L(T : T_o, D)]^p [F(T : T_o, D)]^q \quad (5)$$

where $z = [C]/69$, $h = [H]/6$, and $x = [O]/25$, and the parameters α, β, and γ, T_o, D, p, and q were found to have simple relationships to the feedstock and product defining parameters [H], [O], a, b, and c. The final ASEM formulas that fit the data could then be used to extrapolate or interpolate the XT results to any [H], [O] feedstock and temperature. Fig. 4 gives an overview of the interpolated and extrapolated $Y(T)$ outputs for a selection of products for four representative feedstock along nature's coalification path.

Because hundreds or even thousands of organic products of pyrolysis have been identified in the literature, to go much further, some comprehensive organization of these products is needed. Toward this goal, the CCTL has grouped products into the families shown in Table 4, along with the a, b, and c rules that connect these groups. This list can be subdivided into pure hydrocarbons (i.e., C_aH_b) and the oxygenates (C_aH_bO, $C_aH_bO_2$, CaH_bO_3, etc). Isomers (groups with identical a, b, and c) can differ in

Table 4 Organization of functional groups by family

Families	a	b	c
Paraffins	j	2a+2	0
Olefins	$j+1$	2a	0
Acetylenes	$j+1$	2a−2	0
Aromatics	$5+j$	$4+2j$	0
Polynuclear	$6+4j$	$6+2j$	0
Aldehydes	$j+1$	2a	1
Carbonyls	j	2a	1
Alcohols	j	2a+2	1
Ethers	$j+1$	2a+2	1
Phenols	$5+j$	$4+2j$	1
Formic acids	j	2a	2
Guaiacols	$6+j$	$6+2j$	2
Syringols 1	$7+j$	$8+2j$	3
Syringols 2	$8+j$	$10+2j$	4
Sugars 1	$4+j$	10	5
Sugars 2	$5+j$	$10+2j$	5

a, b, and c are the subscripts in $C_aH_bO_c$, where $j=1, 2, 3\ldots$.

detailed pyrolysis properties and, hence, parameters. We use $j=1$, 2, 3, etc. to denote the first, second, third, etc. members of each group or the carbon number (n). In the CCTL's most recent studies[21–27] of specific feedstock pyrolysis, formulas have been proposed and tested for the dependence of the W, T_0, D_0, p, and q parameters on the carbon number of the product within each group. This makes it possible to compact a very large body of data with simple formulas and a table of parameters.

The case of polyethylene is an example of such a study. It is not shown in Fig. 2a, as it is far removed from the coalification curve, having the position [H]=14.2 on the [O]=0 axis. Without oxygen in the feedstock, the pyrolysis products are much fewer, and the ASEM is much simpler to use than with carbohydrates. Thus, only the first five rows of Table 4 are needed to cover the main functional groups involved in organizing the pyrolysis products of polyethylene.

Fig. 5 gives an ASEM-type summary of the product yields vs temperature based on fits to the experimental data of Mastral et al.[30,31] at five temperatures that were constrained to satisfy approximately mass, carbon, and hydrogen balances. When the parameter systematics are identified, the ASEM representation can be used to estimate the pyrolysis product of polyethylene pyrolysis at any intermediate temperature or at reasonable extrapolated temperatures. The experimental data was available only up to 850°C, but the extrapolations to 1000°C were constrained in detail to conform to mass, carbon, hydrogen, and oxygen balance.

Fig. 5 also shows extrapolations to 6000°C that may be of interest if one goes to very high temperatures—by plasma torch heating, for example. Here, we incorporate a conjecture that at the highest temperatures, carbon and hydrogen emerge among the products at the expense of the C1–C2 compounds, as well as aromatics and PNAs components.

Although we have already found that an ASEM can begin to bring some order and overview into pyrolysis yields, clearly, we have a long way to go. When the time dimension is important, the overall search is for a reasonable function of seven variables: [H], [O], a, b, c, T, and t. Einstein's special relativity dealt with only four variables: x, y, z, and t.

SW–IGCC VS NGCC AND ACE

Before World War II almost every town had its own gas works, mainly using coal, as a feedstock. After World War II cheap NG became available and became a major PES for home heating and cooking, as well as for industrial purposes. In the 1980s factory-produced NGCC became available, and NG became a baseload fuel source for many electric utilities, hastening the drawdown of U.S. domestic supplies. In the past four years NG prices have risen to some three to seven times greater than they were when most of these NGCC facilities were built. Thus, pursuing SWEATT is very timely. For most biomass and plastic feedstock, pyrolysis is substantially equivalent to gasification.

The economic feasibility of using a gasifier in front of a gas-fired system can be examined with simple arithmetic and algebra using an analytical cost estimation (ACE) method.[6–10] Analytical cost estimation takes advantage of the almost-linear relationship shown in many detailed cost analyses of the cost of electricity (COE=Y) vs the cost of fuel (COF=X) for many technologies, i.e.,

$$Y(X) = K + XS \qquad (6)$$

Here, Y is given is in cents/kWh, and X is given in dollars/MMBtu. In Eq. 6, S is the slope of the $Y(X)$ line in cents/kWh/$/MMBtu or 10,000 Btu/kWh. S relates to the net plant heat rate (NPHR) via

$$S = \text{NPHR}/10{,}000 \text{ or efficiency via}$$

$$S = 34.12/\text{Eff} \qquad (7)$$

For modern coal plants $S\sim1$, although supercritical pulverized coal (SCPC) plants are reaching toward 0.9 [6–10]. Essentially, the parameter K=COE if the fuel comes to the utility without cost.

In previous studies,[6–10] we assigned $K_{ng}=2$ as a reasonable zero fuel cost parameter for, say, a 100 MW NGCC system.[32,33] This low number reflects the low capital costs of the factory-produced gas turbines and

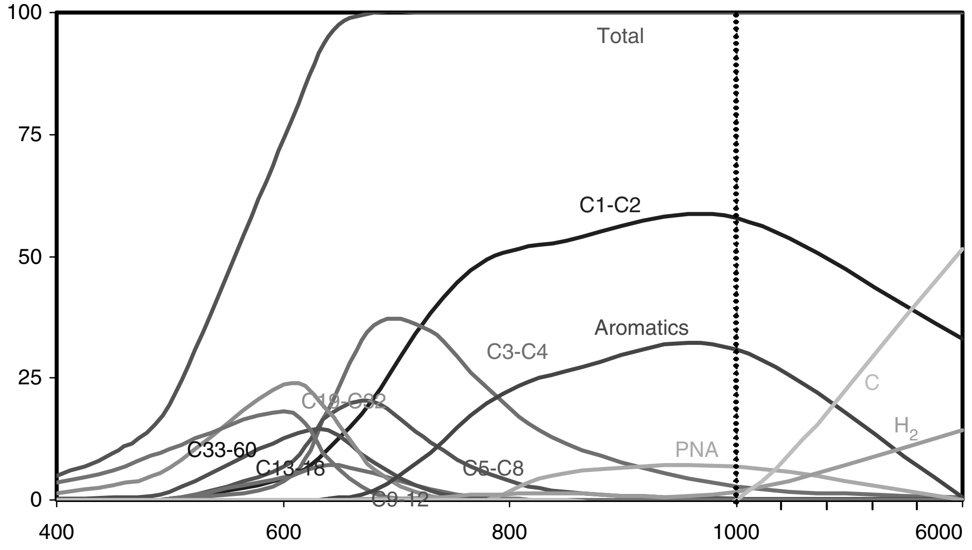


Fig. 5 Yields vs temperature for polyethylene in various hydrocarbon groups.
Source: Adapted from Ref. [23].

steam turbines in NGCC systems reasonable om and en costs. A slope $S_{ng} = 0.7$ is a reasonable assignment reflecting the high efficiency of recent NGCC facilities.

For a SW integrated gasification combined cycle (SW–IGCC) system, K_{sw} generally would be higher than K_{ng} because the capital costs and operating cost must include the gasifier and gas cleanup system. The value of S_{sw} is also higher than S_{ng} because we must first make an SES producer gas, syngas, or pyrogas,which involves some conversion losses. $S_{sw} = 1$ is a reasonable ballpark slope for an up-to-date SW–IGCC system. The X_{sw} for a SW–IGCC system that would compete with a NGCC system at various X_{ng} must satisfy

$$K_{sw} + X_{sw}S_{sw} = K_{ng} + X_{ng}S_{ng} \quad (8)$$

By algebra, it follows that the SW fuel cost X_{sw} that would enable a SWCC system to deliver electricity at the same cost as a NGCC system paying X_{ng} is given by

$$X_{sw} = (K_{ng} - K_{sw})/S_{sw} + X_{ng}(S_{ng}/S_{sw}) \quad (9)$$

In what follow, all X numbers are in dollars/MMBtu, and all Y and K numbers are in cents/kWh. Let us use Eq. 9 with $K_{ng} = 2$, $S_{ng} = 0.7$, $S_{sw} = 1$ and $K_{sw} = 4$ as a reasonable ballpark numbers based on several SWCC analyses.[6–10] Then the first term in Eq. 9 is -2. Now when the $X_{ng} = 2$ to generate SWCC electricity at the same cost, the SW provider must deliver the fuel at a negative price—i.e., pay the tipping fee -0.7. If X_{ng} is near 6, however, as it was in 2004 and in the spring and summer of 2006, the SWCC utility could pay up to 2.4 for the SW fuel. If the X_{ng} is 12, the SWCC facility could pay 7.1 to the SW supplier. This X_{sw} price is much higher than that of coal, the delivered price (X_c) of which these days usually is in the 2–3 range. This simple cost comparison is illustrated in Fig. 6, which shows the opportunities for SWCC systems when NG prices are above, say, \$5/MMBtu. The results are slightly less favorable if the K_{sw} were higher—say, $K = 5$. The conclusion, however, that at high NG prices SWCC electricity becomes competitive with NGCC electricity would be similar. It is conceivable that K_{sw} could be held as low as 2 cents/kWh by retrofitting a NGCC system stranded by high NG prices. In this case, the first term in Eq. 9 vanishes, and the competitive $X_{sw} = (S_{ng}/S_{sw})X_{ng}$. This illustrates the main point that at high NG prices with an ATT system, SW can be a valuable PES. Indeed, this simple algebraic exercise establishes the feasibility of a new paradigm in which SW (mostly biomass but here meaning all solids that are now wasted) becomes a potentially valuable marketable asset.

As described above, the values of K and S are the key factors in determining the COF_{sw} to be used in a SW–IGCC that would be competitive on a COE basis with the COE using a NGCC system at the available COF_{ng}. The ACE method can be extended to the use of SW

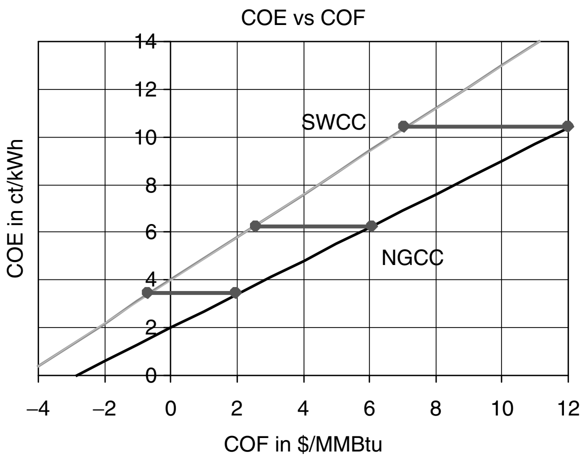


Fig. 6 COE vs COF for SWCC and NGCC at $X_{ng} =$ 2, 6, 12.
Source: Adapted from Ref. [10].

or biomass with other technologies if we can identify the K and S for each technology.

The CCTL has applied the ACE method to a large body of COE-vs-COF calculations on biomass use presented in an Antares Group, Inc. report (AGIR)[34] for several technologies. It is reasonable to apply these results to most of the SW listed in Table 1, particularly in small communities that have recycling programs involving residential separation of waste that would minimize the cost of making RDF.

The technologies investigated in the AGIR when 100 tons per day forest thinning was available include a biomass-integrated gasifier combined cycle (B–IGCC) system, a B-IG simple cycle (B-IGSC) system, a BIG internal combustion (B-IGIC) system, a biomass-gasification–coal cofiring BIG–CC of system, a direct co-firing of biomass and coal in a coal-steam boiler BCoSt, a direct use of biomass in a feedwater heat recovery arrangement (FWHR), direct use of biomass in a stoker fire boiler steam turbine (SFST) system. and direct firing in a combined heat and power plant (CHP) with a steam market at $6/ MMBtu. For each technology, it was possible to approximately represent the tabulated COE vs COF in Eq. 6 and to evaluate K and S. Then, using reasonable analytical forms for $K(P)$ and $S(P)$, one can make COE comparisons at various power levels and fuel costs. The most interesting result of this ACE digest of the massive AGIR tables is the fact that with slight extrapolations to higher power levels, the results in several important cases were opposite those for lower power levels. Because the assumed 100 tons per day of forest residue was rather low for many areas, these changes in conclusions were important.

Thus far, we have focused on the competition between NG-fueled technologies and SW-fueled technologies. Competition of SW-generated electricity with coal-steam-generated electricity appears to be a bigger problem. If one includes the more expensive scrubber cost in the Ks and externality cost in the coal X's,[35] however, the SW–IGCC route should fare well. Coal burning is a major issue in many communities, yet when one projects technology directions around the globe, it is clear that the Gasification Age is returning.[36]

COMPONENTS OF ACE (CACE)

The CCTL is in the process of examining other economic COE-vs-COF analyses to quantify a more detailed formulation of ACE in which K is broken into components $K = K_c + K_{om} + K_{en}$, where c stands for capital costs, om for operating and maintenance costs, and en for environmental costs. At this time, establishing the magnitudes of these components for various technologies and power levels is at the cutting edge of utility economic analyses, and there are large disagreements—particularly on K_{en}. Future fuel costs (X) in the COE term XS in many important cases, however, probably represents the greatest source of uncertainty. Accordingly, it is foolish to belabor estimating K factors with great precision when X can range over wide limits. In this component form of ACE, Eq. 6 is replaced by

$$Y = K_{cr}(P_r/P)^{\alpha} + K_{omr}(P_r/P)^{\beta} + K_{er}(P_r/P)^{\gamma} + XS_r(P_r/P)^{\delta} \qquad (10)$$

where K_{cr}, K_{omr}, K_{er}, and S_r are established on the basis of a detailed analysis at a reference power level P_r, and α, β, γ, and δ are scaling parameters intended to reflect the tendency of per-energy-unit cost to go down as the power goes up (economy of scale). These scaling parameters might be established on the basis of a broad set of studies for each technology.

Table 5 lists CACE parameters extracted from a detailed analysis, "Options for Meeting the Electrical Supply Needs of Gainesville," prepared by ICF Consulting.[35] Here, the final cost of electricity is given in 2003 cents/kWh. The third case, NGCCc, has been added to allow for the contingency that with new offshore drilling or the increased development of liquefied natural gas

Table 5 CACE applied to the ICF consulting study

Tech	Pr	K_0	K_{om}	K_{en}	Xr($/MMBtu)	S_o	COEr(ct/kWh)
NGCCa	220	0.598	0.234	−0.170	6.10	0.68	4.81
NGCCb	220	0.598	0.234	−0.170	11.34	0.68	8.37
NGCCc	220	0.598	0.234	−0.170	4.00	0.68	3.38
SCPC	800	1.491	0.299	0.714	1.91	0.93	4.28
CFB-CB	220	2.531	0.261	0.618	1.41	1.05	4.89
CFB-B	75	2.845	0.261	0.039	1.67	1.39	5.47
IGCC	220	2.2	0.196	0.407	1.41	0.86	4.02

Solid–Sus

(LNG) importing capabilities, NG might return to the $4/MMBtu level of 2003.

The final column shows that at the reference power levels without the NGCCs case, the IGCC scores the lowest COE, as the ICF report (ICFR) concluded. The value of the ACE analysis is that with a bit of algebra, anyone can easily consider other fuel cost projections and other power levels (with assumed values of α, β, γ, and δ). Based on previous CCTL exploratory work and economy-of-scale investigations, the author estimates that for costly field-erected facilities, $\alpha = \beta = 0.3$ are reasonable choices. With factory fabrication of jet and steam turbines, however, these parameters may not follow the usual economy-of-scale pattern and be somewhat smaller in magnitude. Assigning a value for γ is a wide-open question because environmental costs and the methods of incorporating them into the cost of electricity are highly debated issues.[36] Reasonable values for δ are also somewhat difficult to find. For NGCCs, the author tentatively assigns close to zero or a very small value (~ 0.1), perhaps because the development of highly efficient aeroderivative turbines has proceeded on a wide range of power levels.

In concluding this section, it should be clear that the age of making gas has returned and that the time has come to develop a national gas strategy[37] to facilitate the earliest implementation of new gasification systems.

BIOFUELS

Biochemical Conversion by Fermentation

Fermentation, a major form of biochemical conversion, uses bacteria in the presence of oxygen to break down biodegradable organic material into liquid fuels such as ethanol. The ethanol thrust is an extension of the commercial beer, wine, and alcohol industries' processing of sugar- and starch-based feedstock such as corn and sugarcane. An important development,[38] however, has extended these capabilities to cellulose, greatly expanding the mass of biomass that can be transformed by the fermentation route. Ethanol lends itself to conventional automotive liquid fuel storage, although at only 0.6 times the energy density of diesel or gasoline. As Brazil has demonstrated, an automotive fleet can be largely fueled by ethanol. An aircraft fleet will require a higher-energy-density fuel, however.

Biochemical Conversion by Anaerobic Digestion

In anaerobic digestion, bacteria convert biodegradable waste to methane gas, a technology that is an extension of the phenomenon of flatulence of animals.[39] Landfill gas is a product of anaerobic digestion. Although a good-quality gas can be achieved by anaerobic digestion, escaping methane can be a problem, because one molecule of methane is some 20 times more damaging as a greenhouse gas than carbon.

At this time, neither the aerobic nor the anaerobic biochemical conversion is capable of processing lignin (approximately 25 wt% of plant matter) or any plastics except biodegradable plastics. The main disadvantage of bioconversion is that the reaction times are weeks, whereas thermochemical conversion can take place in minutes. Thus, the volume required for large-scale biochemical processing is very much larger than for ATT processing. The fact that the residue from biochemical conversion can be a good feedstock for thermochemical processes[1] suggests that bioconversion and thermal conversion should work together. Estimates of SW from alcohol and methane production are included in Table 1.

Bio-Oils

The esters of vegetable oils are renewable alternative fuels that potentially can serve as direct replacements for diesel fuels in compressed ignition engines (CIE).[39] Oils from soybeans, sunflower seeds, safflower seeds, cotton seeds, peanuts, and rapeseeds, as well as used oil from restaurants, are under considerable investigation as replacements for diesel. Waste from bio-oil programs have been included in Table 1 because only the seeds of the plants are used for bio-oil crops; the rest of the plant becomes SW amenable to serving as an input of a SWEATT program.

RECYCLING AND SWEATT

Although our confrontational society has a tendency to view waste-to-energy as a threat to recycling programs, the opposite may be true. Recycling programs in a community can serve to sort the various components of municipal or institutional SW into categories that lend themselves to maximizing the return on these components. If, for example, newspaper at a given time has no recycling market but must be disposed of at a cost, it could be used as high-energy dry feedstock for ATTs. The same is true for plastics recycling. Thus, the marketplace would be decisive as to whether to recycle via the materials route or the energy route. A recycling community should be able to go the SW–IGCC route with less capital costs than one that does not have waste separation at the source.

The advantages of a biomass alliance with natural gas (BANG) have been described previously.[6–10] Gasification systems that mainly use cellusosic (biomass) inputs produce a low- or medium-heating-value fuel that will result in the derating of a NG designed turbine-generator. By coutilizing the biomass pyrogas with NG, one can ensure that the input energy requirement matches the output needs at least until the maximum rating of the generator is required. At that point, the firing could be entirely on NG. In a solid waste alliance with natural gas (SWANG), an additional option becomes available when the SW comes from a recycling community. Then the utility might prepare and store high-energy plastics for increased use during times of high electricity demand as a means of following peak loads without calling on the full use of NG.

ATT FOR LIQUID FUEL PRODUCTION

Pyrolysis/gasification technologies followed by gas cleanup can greatly reduce emissions of pollutants such as NO_x and SO_x, as well as toxics such as mercury and arsenic. Advanced thermal technologies can treat nearly the entire organic fraction of MSW and, in general, can treat a more heterogeneous feedstock, including high-energy-content plastics.[11] Although this paper focuses on gaseous fuel generation, ATTs for liquid fuel (condensable gas) production are closely related. Considerable research and development work is under way on distillation technologies to refine such liquid fuels for transportation applications, adding a major driver for the ATT route.

It should be noted that Table 1 does not list oil shale or tar sands in the United States that could substantially increase the available SW tonnage that could be used to address our need for transportation fuels. A 2005 Rand study[40] shows that with in-situ thermal treatment, domestic oil shales could substantially lower our oil import problem. Another route would be to convert our coal to liquid fuels, as South Africa has done for many years. A third route would involve the use of methane hydrates to produce methane for use in NG-fueled vehicles.

CONCLUSIONS

The main conclusion of this paper is that the United States has very large sustainable supplies of now-wasted solids that have an annual energy potential comparable to that of our current use of coal and also of NG. With ATT, this SW could, in the near term, multiply its contribution to our national energy supply by a factor of about 10. Robust technology that can handle MSW or RDF also should be able to handle almost any of the categories listed in Table 1. Agricultural and forestry residues are two of the major SW supply components in this list, and many of the other materials are greenhouse-neutral plant matter.

By utilizing thermochemical processes to convert the lignin and plastic content of SW and biochemical residues, we could get much closer to goal of zero SW. Solid waste to energy by advanced thermal technologies also would further reduce the final volume of the waste, and practically all contaminants can be destroyed by high temperatures. Thus, cooperation between biochemical and thermochemical programs would clearly be in the national interest.

Japan, a country with an outstanding sustainability record, is the global leader in SWEATT. With more than 60 pyrolysis and thermal gasification systems now in operation, Japan has established the technical and environmental feasibility of these systems. This should allay the concerns of environmentalists and risk-averse utility decision makers in the United States.

In final summary, our main conclusions are

- The United States is excessively reliant on imported oil (60%) for its liquid fuels.
- The United States is increasingly reliant on imported NG fuel (now $>15\%$).
- The United States is well endowed with solid fuels: wasted solids, coal, and oil shale.
- The organic matter in SW can be converted to useful gaseous or liquid fuels.
- In most cases, ATT provides the fastest and most efficient conversion method.
- Co-use of domestic fuels such as SW and NG can overcome some problems.
- Solid waste to energy by advanced thermal technologies generates lower emissions than combustion waste-to-energy systems.
- Thermal conversion of solid fuels to gaseous and liquids fuels has a long history.
- Utilities are experienced with high-temperature processes in the production of steam.
- Advanced thermal technologies (PABC, POBC, and PYRO) are extensions of high-temperature steam making.
- Conversion to gaseous fuels is essential for SW powering of fuel cells.
- There are many environmental benefits attendant to SWEATT.
- Many areas of engineering research will be needed to optimize SWEATT.
- Cooperation of stakeholders would accelerate the implementation of ATT.
- Conservation and SWEATT together is the fastest realistic path to zero waste.

ACKNOWLEDGMENTS

The author would like to thank his many coauthors in this 30-year quest for alternatives to oil; their contributions are incorporated into this article.

Solid–Sus

Glossary

ABPC: Air Blown Partial Combustion
ACE: Analytical Cost Estimation
AGIR: Antares Group Inc. Report
Ar: Aromatics
ASEM: Analytical Semiempirical Model
ATT: Advanced Thermal Technologies
BTU: British Thermal Units
CACE: Component Analytic Cost Estimation
CCTL: Clean Combustion Technologies Laboratory
CHP: Combined Heat And Power
CIE: Compressed Ignition Engines
DANSF: Dry Ash, Nitrogen And Sulfur Free
EU: European Union
FC: Fixed Carbon
GT: Gas Turbines
HHV: Higher Heating Values
HRSG: Heat Recovery Steam Generator
ICE: Internal Combustion Engines
IGCC: Integrated Gasifier Combined Cycle
MSW: Municipal Solid Waste
NGCC: Natural Gas-Fired Combined Cycle
NPHR: Net Plant Heat Rate
OBPC: Oxygen Blown Partial Combustion
OFC: Omnivorous Feedstock Converter
PES: Primary Energy Supplies
PNA: Polynuclear Aromatics
PYRO: Pyrolysis Systems
QES: Quaternary Energy Supply
quads: Quadrillion BTUs
RDF: Refuse Derived Fuels
SES: Secondary Energy Supplies
SW: Solid Waste
SWANG: Solid Waste Alliance with Natural Gas
SWEATT: Solid Waste To Energy By Advanced Thermal Technologiess
TES: Tertiary Energy Supply
TPES: Total Primary Energy Supply
VT: Volatiles
WEC: Waste to Energy Conversion

REFERENCES

1. Green, A., Ed. *Alternative to Oil, Burning Coal with Gas*, University Presses of Florida: Gainesville, FL, 1981.
2. Green, A. et al. Coal-water-gas: an all American fuel for oil boilers. Proceedings of the Eleventh International Conference on Slurry Technology, Hilton Head, SC, 1986.
3. Green, A. Ed. Solid fuel conversion for the transportation sector, In FACT. Vol. 12. Proceedings of Special Session at International Joint Power Generation Conference, San Diego, CA; ASME: New York, 1991.
4. Green, A. A Green Alliance of Biomass and Coal (GABC). Appendix F. *National Coal Council Report; May 2002*, Proceedings of the 27th Clearwater Conference, March 2003.
5. Green, A.; Feng, J. A Green Alliance of Biomass and Natural Gas for a Utility Services Total Emission Reduction (GANGBUSTER). Final Report to School of Natural Resources and the Environment, 2003.
6. Green, A.; Hughes, E.; Kandiyoti, R. Proceedings of the first international conference on co-utilization of domestic fuels. Int. J. Power Energy Syst. **2004**, *24* (3).
7. Green, A.; Smith, W.; Hermansen-Baez, A. et al. Multi-disciplinary Academic Demonstration of a Biomass Alliance with Natural Gas (MADBANG). Proceedings of the International Conference on Engineering Education, Gainesville, FL, October 2004.
8. Green, A.; Swansong, G.; Najafi, F. Co-utilization of domestic fuels biomass gas/natural gas. In *GT2004-54194*; IGTI: Vienna, Austria, 2004; [June 14–17].
9. Green, A.; Feng, J. Assessment of technologies for biomass conversion to electricity at the wildland urban interface. Proceedings of ASME Turbo Expo, Reno–Tahoe, NV, 2005.
10. Green, A.; Bell, S. Pyrolysis in waste to energy conversion. Proceedings of the NAWTEC 14 Conference, Tampa, FL, May 2006.
11. California Integrated Waste Management Board. Conversion Technologies Report to the Legislature [draft], Available at: www.ciwmb.ca.gov/Organics/Conversion/Events.
12. Energy Information Administration. *January 2006 Monthly Energy Review*, DOE/EIA-0035(2006/01); Available at: www.eia.doe.gov/emeu/mer/contents.html (accessed on January 2006).
13. Perlack, R.; Stokes, B.; Erbach, D. Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply, ORNL/TM-2005/66. Oak Ridge National Laboratory: Oak Ridge, TN, 2005.
14. Stultz, S., Kitto, J., Eds. *Steam* 40th Ed.; Babcock and Wilcox: Barberton, OH, 1992.
15. Green, A.; Peres, S.; Mullin, J. et al. Co-gasification of domestic fuels. Proceedings of the IJPGC; ASME: Minneapolis, MN, 1996.
16. Green, A.; Zanardi, M.; Mullin, J. Biomass Bioenergy **1997**, *13*, 15–24.
17. Green, A.; Zanardi, M. Int. J. Quantum Chem. **1998**, *66*, 219–227.
18. Green, A.; Mullin, J. J. Eng. Gas Turbines Power **1999**, *121*, 1–7.
19. Green, A.; Mullin, J.; Schaefer; G. et al. Life Support Applications of TCM-FC Technology. Presented at: 31st ICES Conference, Orlando, FL, July 2001.
20. Green, A.; Venkatachalam, P.; Sankar, M.S. Feedstock Blending of Domestic Fuels in Gasifier/Liquifiers. Presented at: TURBO EXPO 2001, Amsterdam, 2001.
21. Green, A.; Chaube, R. Pyrolysis Systematics for Co-utilization Applications, Presented at: TURBO EXPO 2003, Atlanta, GA, June 2003.
22. Green, A.; Chaube, R.A. Int. J. Power Energy Syst. **2004**, *24* (3), 215–223.
23. Green, A.; Sadrameli, S.M. Analytical representations of experimental polyethylene pyrolysis yields. J. Anal. Appl. Pyrolysis **2004**, *72*, 329–335.
24. Sadrameli, S.; Green, A. J. Anal. Appl. Pyrolysis **2005**, *73*, 305–313.
25. Green, A.; Feng, J. Systematics of corn stover pyrolysis yields and comparisons of analytical and kinetic representations. J. Anal. Appl. Pyrolysis **2006**, *76*, 60–69.

Solid–Sus

26. Feng, J.; Green, A. Peat Pyrolysis and the Analytical Semi-Empirical Model. J. Energy Sources, **2006**. In press.
27. Feng, J.; YuHong, Q.; Green, A. Analytical model of corn cob pyroprobe-FTIR data. Biomass Bioenergy **2006**, *20*, 486–492.
28. Gaur, S.; Reed, T. *Thermal Data for Natural and Synthetic Fuels*; Marcel Dekker: New York, 1998.
29. Xu, W.C.; Tomita, A. Effects of coal type on the flash pyrolysis of various coals. Fuel **1987**, *66*, 627–631. 632–636.
30. Mastral, F.; Esperanza, J.; Garcia, E. et al. J. Anal. Appl. Pyrolysis **2002**, *63*, 1–15.
31. Mastral, F.; Esperanza, J.; Berruco, E. et al. J. Anal. Appl. Pyrolysis **2003**, *70*, 1–17.
32. Liscinsky, D.; Robson, R.; Foyt, A. et al. Advanced technology biomass-fueled combined cycle. Proceedings of ASME Turbo Expo, ASME: Atlanta, GA, 2003.
33. Phillips, B.; Hassett, S. Technical and Economic Evaluation of a 79 MWe (Emery) Biomass, Presented at: IGCC Gasification Technologies Conference, San Francisco, CA, 2003.
34. Antares Group, Inc. *Assessment of Power Production at Rural Utilities Using Forest Thinnings and Commercially Available Biomass Power Technologies*; Antares Group, Inc.: Landover, MD, 2003.
35. ICF Consulting Options for Meeting the Electrical Supply Needs of Gainesville report submitted March 6, 2006.
36. Roth, I.F.; Ambs, L.L. Incorporation externalities into a full cost approach to electric power generation life-cycle costing. Energy **2004**, *29* (12–15), 2125–2144.
37. Rosenberg, W.; Walker, M.; Alpern, D. *National Gas Strategy*; Harvard University Press: Cambridge, MA, 2005.
38. Ingram, L.; Alterthum, F.; Conway, T. Ethanol production by *Escherichia coli* strains co-expressing zymomonas pdc and adh genes Us Patent No. 5,000,000, 1991.
39. Wilkie, A.C.; Smith, P.H.; Bordeaux, F.M. An economical bioreactor for evaluating biogas potential of particulate biomass. Bioresour. Technol. **2004**, *92* (1), 103–109.
40. Bartis, J.; LaTourrette, T. et al. *Oil Shale Development in the United States: Prospects and Policy Issues*; National Energy Technology Laboratory: Pittsburgh, PA, 2005.

Space Heating

James P. Riordan
Energy Department, DMJM+Harris/AECOM, New York, U.S.A.

Abstract
Nothing seems more welcoming than the coziness of a blazing fireplace on a cold winter night. The technological developments leading up to this inviting scene are the result of over 2000 years of developments. The technological advances within the past 100 years and in our future will shape new industries, create better energy awareness, and mark changes in our lifestyles.

INTRODUCTION

Since early human history, mankind has used fire for protection, lighting, cooking, and generating warmth in cold climates. Fire, utilizing many miscellaneous fuels, is still the main heat source used for space heating worldwide. Over time, the fuels and methodology have varied considerably. Many sources of fuel are available worldwide. Some of the sources that are utilized are wood, coal, oil products, natural gas, geothermal, biomass, peat, solar, wind, and hydro-produced electric energy. The recent use of efficient building design using high *R*-value insulation and updated technologies has allowed us to redefine industry standards for space heating.

This entry will look at many historical methodologies used for space heating and the improvements made to the processes along the way. Next, it will investigate the current technologies in use, and how they relate to existing energy conservation and indoor air-quality standards. Last, the need for more energy-efficient equipment and better public energy awareness will be addressed.

HISTORICAL INFLUENCES

Thousands of years ago, as mankind explored and populated the continents, heat from fire became an invaluable commodity. It allowed communities to cook, protect themselves, and provide light, and it was also used for the processing of metals, leading to the evolution of tools. As technology advanced, these tools were utilized to build improved structures for habitation and public use. Early habitats used fire pits to radiate heat by a combination of convection, conduction, and radiation in an attempt to keep the inhabitants relatively comfortable. The problems incurred with this type of system were the labor involved in supplying the wood fuel and how to get rid of the smoke. Many early huts were built with holes in the roofs to allow the smoke out, but this also allowed any inclement weather to come in. The dangers involved with this method are not only the risk of fire due to the building materials, but also the danger of lingering smoke that could cause suffocation.

The Romans developed a method for heating buildings that is evident from excavations of ancient villas and bathhouses. This Roman system of generating radiant heat is called a hypocaust. Modern variations of this radiant heat-type system are still in use today. One of these villas, at Newport on the Isle of Wight in Great Britain, is a good example of such a system. The structure's finished floor is supported by numerous tile columns. Under the finished floor is a chamber or crawl space. The crawl space is usually no more than 2 ft high. Apparently, this was found to be the most efficient design. At one end of the building is a furnace or a fire pit where wood is burned, generating hot air. The pit is usually large enough to allow for a reasonably sized fire while also allowing airflow around the fire and into the crawl-space chamber. The walls of the structure include flues, which are scattered around the structure and made of box-shaped tiles stacked one upon another. These flues ascend from the crawl space through the walls, allowing the hot exhaust gases to escape above the rooftop level. This rush of hot air through the crawl space and up the stacks generates what we know today as a stack draft. This negative pressure pulls more heated air from the furnace through the crawl space, heating the floor. The walls are also heated by these multiple hot stacks. Naturally, the floors and walls closest to the furnace

Keywords: ASHRAE: American Society of Heating, Refrigerating and Air-Conditioning Engineers; ASME: American Society of Mechanical Engineers; BTU: British thermal unit—energy required to raise 1 lb of water 1°F; Conduction: The transmission of energy through a nonmoving medium; Convection: Heat transfer through liquid or gas by circulation; Energy Policy Act 1992: Federal legislation adapted enacting many energy standards; Geothermal: Relating to the internal heat of the earth; Heating Degree Day: Traditional procedure for calculating fuel consumption; Radiation: Energy radiated or transmitted by waves or rays; R-Factor: Insulating factor=h-ft^2-F/BTU; Space Heating: The study and science of heating habitable structures; Stack Draft: Negative stack pressure caused by evacuation of hot gas; *U*-Factor: Inverse of *R*-factor (1/R); *U*=BTU/h-ft^2-F.

Fig. 1 Roman bath (Courtesy Bath and North East Somerset Council, England).
Source: From http://www.romanbaths.co.uk/

are the warmest; those farther away do not receive as much heat. This was an early form of radiant heating.[1]

In today's culture, bathing is a private activity. In Roman culture, communal bathing and public bathhouses were widespread leisure activities and fundamental staples of Romans' lives. Many bathhouses were privately owned (called balneae), but there were also many public bathhouses (called thermae) (see Fig. 1). Public bathhouses were accessible by all levels of society—by both men and women—for a trivial fee. The remains of the Stabian Baths in Pompeii are a good example of a public bathhouse. Many bathhouses were lavish in their construction and decoration. Usually, they allowed woman to use them in the morning hours, while men were allowed entry from about 2:00 P.M. until sunset. These bathhouses also provided facilities for sports and recreation, and acted as a community center for cultural and intellectual interaction. Many included gardens, libraries, entertainment areas, and lecture halls.

A typical visit usually started in the warm room (tepidarium), which had heated walls and floors. From there, the client normally would proceed to the hot bath (caldarium) for a leisurely bath. This hot bath normally would be located closest to the furnace. After this, the client would proceed to the cold room (frigidarium) and soak in a cold pool of water. Many bathhouses offered steam rooms and dry heat similar to a sauna (laconicum). Needless to say, these were widely used public buildings handling large occupancies.[2]

The thermal demand for a private-residence bathhouse would be moderate. The thermal demand for a public bathhouse must have been tremendous. The furnace fires would require regular attention from many servants, and obviously, this was extremely labor intensive. The furnace(s) would require frequent addition of quantities of wood and removal of the hot ashes. The wood burned would be mainly small branches up to about 3 in. in diameter and up to 3 ft long. Logs were not effective, as they burned too slowly and blocked the airflow into the crawl-space chamber. The height of the flame was also restricted to about half the height of the chamber to allow for proper airflow.[3]

HEARTHS AND FIREPLACES

During the Middle Ages, only slight improvements were made in space heating. The open hearth preceded the fireplace, and saw wide use in Saxon times and later. The hearth was usually bordered by stone or tile, and smoke rose through a rooftop louver. A "couvre-feu," or fire cover, made of tile or china was placed on the hearth at night to reduce the fire risk. The disadvantages of an open hearth were that the warmest locations in the building were closest to the fire. Smoke also became a fire and health issue. The fireplace, which provided heat both directly and by radiation, saw wide use later during this period. The stones in the back of the hearth and the walls opposite the

Fig. 2 Uncovered medieval fireplace (Courtesy of Chichester District Council, England). Source: From http://www.chichestertoday.co.uk/viewarticle2….

fireplace were usually made extra thick to radiate stored heat after the fire had burned low.

When multiple-story structures began utilizing fireplaces, the danger of a fire became critical, as most roofs were made of wood. Hearths and fireplaces began to be located close to a perimeter wall where the exhaust smoke could be funneled into the wall, allowing it to vent properly.[4] Near the end of the 12th century, fireplaces began to be constructed with projecting hoods to control the smoke, allowing for a shallow recess. The flues were routed through the walls to discharge above the rooftop level (see Fig. 2).

During this period, space heating was still dependent on the capacity of the household and its ability to keep the fires fed with fuel. Many ornate beds of this age included feather mattresses, fur coverlets, and curtains to reduce drafts. Bed curtains were commonly used into the early 20th century to reduce drafts and maintain personal warmth.

Three major improvements to fireplaces occurred in the 1600s and 1700s. In 1678, Prince Rupert, Duke of Bavaria, invented a system of hinged baffles in an attempt to extract more heat from the fire; it had limited success and was still prone to smoking. Later, Benjamin Franklin produced a design (the Franklin stove; see Fig. 3), which radiated heat but required the exhaust gases to travel up and over a hollow riser extracting convection-heated air. This was somewhat more successful and was later improved upon by David Rittenhouse, one of Franklin's contemporaries. In the late 1790s, Count Rumford came up with a design incorporating a continuously variable flue damper and a sloping fireback. This sloping fireback provides a larger surface area, allowing greater heat radiation. This design is the basis of many fireplaces used today.[5]

A very efficient fireplace design used in many northern climates is the Finnish Masonry Stove (see Fig. 4). The Finnish stove is a more complicated design using a long

Fig. 3 Franklin stove (Courtesy of The Franklin Institute). Source: From http://en.wikipedia.org/wiki/Franklin_stove

Fig. 4 Finnish stove (Courtesy of Peter Moore Masonry, Inc.). Source: From http://www.vtbrickoven.com/masonry/masonry.html

horizontal flue between the stove section and the flue. Convection-heated air is directed around the masonry sections and into nearby rooms. It is quoted that these stoves capture up to 85% of the BTUs burned in the combustion process for radiated heat. Conventional wood stoves capture about 45%–60%.[6]

Odd fact: up until 1900, the Danish war office was heated by a single furnace in the basement of the building. The furnace was used to heat cannonballs. These cannonballs were carried to every office and placed red hot into a metal bin in the fireplace alcoves to radiate heat.[7]

PRESSURE VESSEL DEVELOPMENT

The capability to heat water and generate steam has been known since ancient times. Its viable use has been seen only since the Industrial Revolution. In 1678, Thomas Savery patented the Savery Pump, which used steam to draw water out of mines.[8] It worked only to a depth of 80 ft, and boiler explosions were common due to lack of advancement in pressure-vessel design. Thomas Newcomen vastly improved upon this design with his Newcomen Atmospheric Engine—which, although inefficient, was used for about 60 years.[9] James Watt and his partner, Matthew Boulton, later improved on the process with the use of a separate condenser. The work done by the Boulton-Watt engines was not by steam pushing a piston but by condensing steam creating a vacuum. They later improved the engine by utilizing the steam to push a piston.[10] The boilers used for this were atmospheric pressure boilers. The technology developments required to drive these engines also led to the development of improved boiler designs.

Using boilers as a heat source is commonplace today. In the 18th and early 19th centuries, boilers were temperamental, and catastrophic explosions frequently occurred. The potential applications for steam power seemed unlimited, but the method of safely harnessing and controlling this energy was not sufficiently developed. Early steamships and locomotives were plagued with pressure limitations to maintain safe operation. In 1865, the Mississippi River steamship Sultana exploded, taking over 1200 lives.[11] At this historical point, there were no boiler codes in use anywhere in the United States or elsewhere. In March 1905, in Brockton, Massachusetts, the Brockton shoe-factory boiler exploded. This catastrophic explosion resulted in 58 deaths and 117 injuries. Massachusetts immediately set about to form a Board of Boiler Rules whose charge was to write a boiler law for the state. This state law was published in 1908.

By 1911, a total of 10 states and 19 metropolitan areas had enacted their own separate boiler laws. The requirements of these individual laws varied greatly. Essentially, a boiler built in one state could not be used in another state due to the differences in the individual codes. In 1911, the American Society of Mechanical Engineers (ASME) appointed a committee to come up with a standard boiler code that could be accepted by all states. The intent of this committee in establishing this code was to protect the public and also to reduce the clutter of miscellaneous state regulations. This was done by bringing boiler and pressure-vessel regulations under one national code. The first ASME Boiler and Pressure Codes were published in 1914. Although accepted as a national code within the United States, it is commonly used by many countries as an international standard.[12]

Since 1914, boiler development has consisted primarily of fire-tube and water-tube boilers. Most early boilers, including those that powered steam locomotives, were fire-tube boilers. A fire-tube boiler is essentially a tank chamber with internal tubes that are immersed in water (see Fig. 5). Hot combustion gases flow through these tubes, boiling the surrounding water. The produced steam accumulates above the water level and is used for the

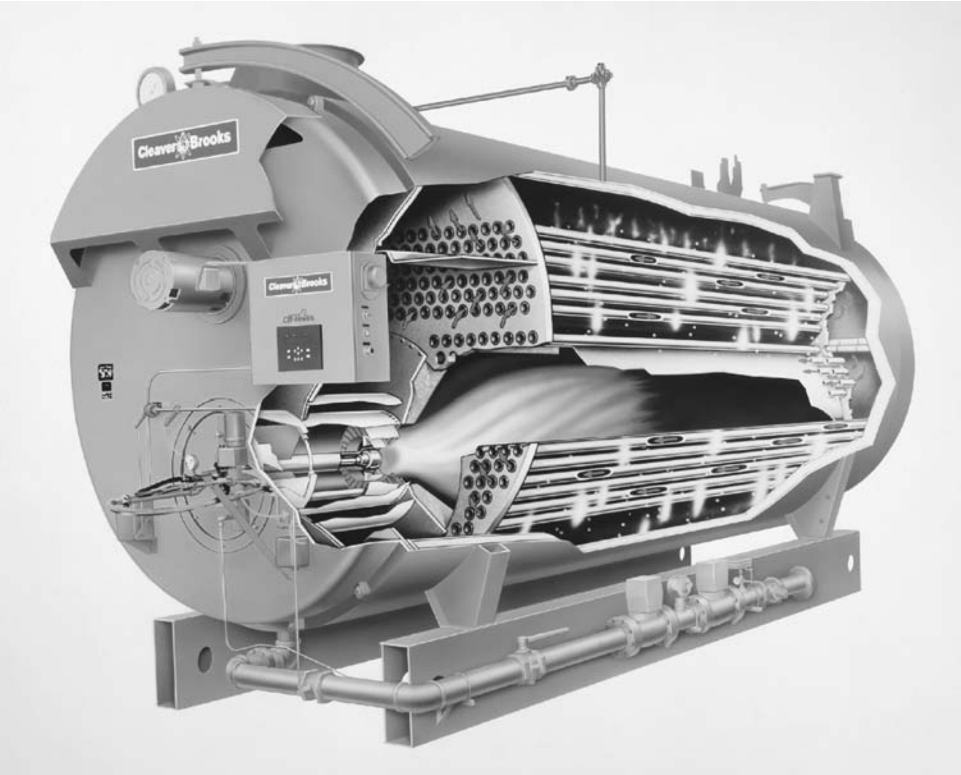


Fig. 5 Fire-tube boiler (Courtesy of Cleaver-Brooks).
Source: From http://www.elizabethtowngas.com/BusinessCustomers.CommercialBoilers.aspx

heating process. Steam traps in the system allow the steam to condense by losing heat to the surrounding air via radiators. Then this condensate is gravity fed or pumped back to the boiler. A water-tube boiler is essentially a tank with internal tubes (see Fig. 6). The tubes contain water, and the hot exhaust gases are passed around the tubes, producing steam. The steam is collected in a steam drum at the top of the boiler for process heat. The condensate is returned the same way as in the fire-tube system.[13]

HEATING SYSTEMS DESCRIPTION

Steam and hot water (hydronic) heating systems utilize different piping arrangements. Steam systems allow the steam to retain high heating values per unit of mass. This is why many large complexes—such as universities, hospitals, and manufacturing facilities—have a central heating facility that utilizes a steam-type system. This system configuration is more complicated, but it alleviates the need for individual building boilers. Another system commonly seen in larger applications is a high-temperature hot water system. This type of system heats water well above its normal boiling point but maintains a liquid state by sustaining high system pressure. Standard hydronic or hot water heating systems are much less complicated and tend to be more centralized. Due to the relatively low temperatures and pressures involved, hydronic systems are much less costly to maintain and are more commonly used in small-building and residential applications.

FUELS

The fuel used for boiler applications has varied widely. Early boilers and industrial drive engines consumed primarily coal and wood. During the 18th century through the early 20th century, these fuels were plentiful and a relatively inexpensive fuel source. In the late 19th century and early 20th century, with the oncoming of improved transportation, petroleum fuels and natural gas took over as the fuels of choice. They were inexpensive, readily available, and easy to store and transport. Wood and coal are still used to this day in many areas of the country. Many metropolitan areas still have remnants of old underground gas systems underneath their streets. Although no longer in use, some of these conduits were made of wood and were used to supply different forms of gas for street lighting and other applications. The recent limitations and petroleum price increases imposed on this country have caused a sudden emergence of the use of high-efficiency wood and coal stoves in residential heating applications. Coal is still used in many large industrial applications due to its availability, low cost, and high fuel-heating values. Of course, current emissions standards

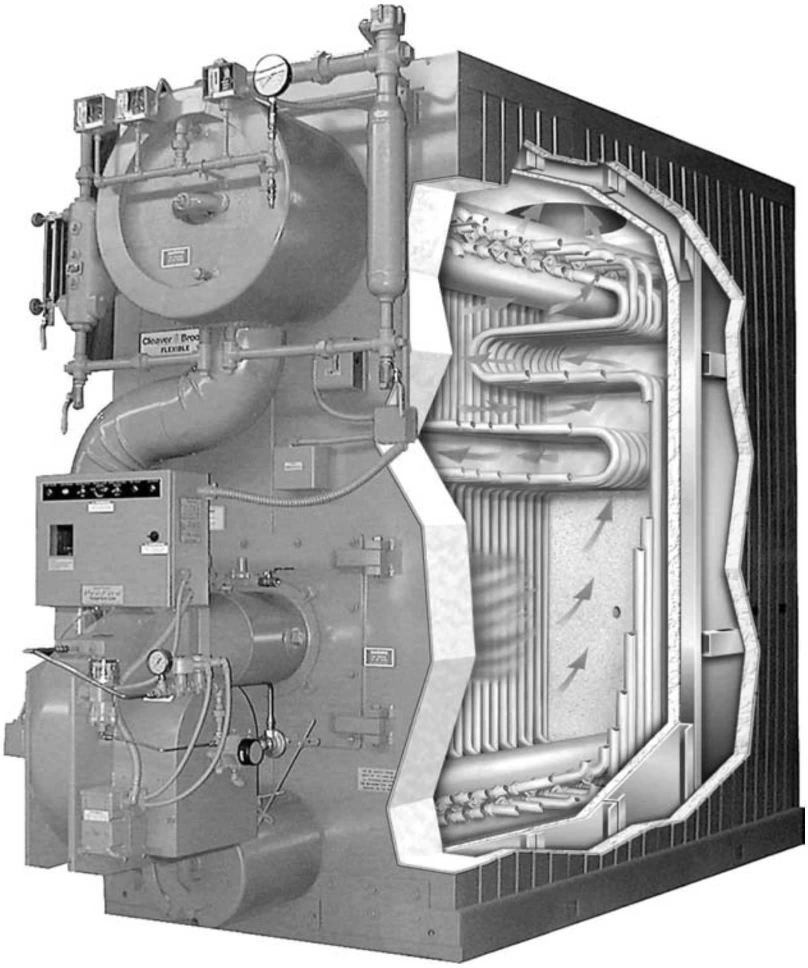


Fig. 6 Water-tube boiler (Courtesy of Cleaver-Brooks).
Source: From http://www.elizabethtowngas.com/BusinessCustomers.CommercialBoilers.aspx

require that special emissions monitoring and treatment remain in effect.

HABITAT DEVELOPMENT

A discussion of space heating would not be complete without reviewing the development of habitat structures and the improvements in the materials of their construction.

Early habitats were constructed primarily of earth, wood, or stone. The roofs were thatched straw, with large openings to allow hearth-fire smoke to escape. The doorways and window openings were rough openings that allowed occupants and light to enter. These doors and window openings may have been covered with animal hides or wood. The structure's main function was to offer some level of protection from predators and the elements. Energy efficiency was not primary on the minds of the occupants. The actual building materials acted as the rough insulation. Drafts causing heat losses were required to allow fire smoke to escape. In moderate climates, the need for protection from the ambient weather would be low to moderate. In the colder latitudes, this became more of a priority. The level of heating required was supplied by maintaining a hearth fire along with heavier clothing, as the climate dictated. The majority of the structures are considered primitive by today's standards.

Later, during the Roman and Middle Ages periods, doors, windows, and shutters came into wider use; these devices reduced air infiltration but did not alleviate it. Glass windows were expensive commodities that were handmade for each application. They saw use more for decorative purposes in wealthier households and churches than for heating efficiency. Many windows leaked and required shutters just to keep the elements out.

In many northern latitudes and during the American colonization period, many log huts or cabins were built. This is only natural, with wood being both plentiful and available. The use of these heavy wood timbers reduced much of the need for thermal insulation. Because wood is fibrous, it acts as a natural insulating thermal barrier.

The openings between the rough-hewn timbers required filling with soil or clay to eliminate drafts, of course. Again, window development did not proceed much past earlier times, and although expensive, it was more readily available. Again, shutters were commonly used for protection from both the elements and predators.

INSULATION DEVELOPMENT

It is only within the past 100 years that the benefits of building insulation have been realized from the standpoint of building comfort and fuel-cost savings. The ideology of efficient insulation practice is to suppress the flow of heat energy from warm areas to cold areas as much as is technically and economically feasible within the guidelines of current air-quality standards. Under normal conditions and in accordance with heat-transfer practices, heat energy will migrate from areas of higher energy (warmer) to areas of lower energy (colder). The rate at which this happens is defined by the thermal conductivity or *U*-factor of the material. The *U*-factor is the inverse of the *R*-factor, or insulating value of the material, in question. Insulating materials that saw use in early structures essentially consisted of the building's structural materials. Later, as frame construction developed and the use of insulation began to develop, whatever would stop the flow of heat was used. This would come to include mineral wool, fiberglass, Styrofoam, and Mylar/aluminum-foil boards.

Insulators can be organic, inorganic or fibrous, cellular, reflective, rigid, soft, or granular. New residences may include many types for different applications. Each type is suited for a specific task, and each type comes in many forms: batts, blankets, loose fill, spray foam, rigid panels, and radiant barriers. Vapor shielding utilized in many homes consisted of single or multiple layers of tar paper; this has steadily been replaced by fiberglass-type blankets. These fabrics have become relatively inexpensive and easy to install, and they essentially eliminate air infiltration where they are used.

Many federal agencies and trade associations have enacted regulation to ensure public safety relating to the use of insulation products. Some of these agencies are the Consumer Product Safety Commission (CPSC), the American Society of Testing Materials (ASTM), and the Federal Trade Commission (FTC). Their regulations and findings are published and easily available for public review.

SPACE HEATING CALCULATION METHODOLOGY

The calculations used to determine acceptable heating comfort standards are in accordance with American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) and local codes, and include cooling requirements and special building applications. These will not be discussed here. This section will focus primarily on heating calculations.

Upon determination of the location of the structure, the ASHRAE heating and wind design conditions table needs to be reviewed. The heating dry-bulb temperature columns in this table are shown in 99.6% and 99% columns.[14] What this indicates, for example, is that 99.6% of the time, the temperature will remain above the listed low temperature. Depending on local codes, this would be your low outside ambient air design temperature. At some points, the temperature could fall below that point, but statistically, this time would be minimal. Considering the desire to maintain an indoor air temperature of 70°F, the difference between the two is the maximum design temperature differential (Δt). The structure's design and material selection should use energy-efficient practices. The materials specified should have high insulating values (*R*-values) and, thus, low heat-loss coefficients (*U*-values). Utilizing the design Δt and the *U*-factors, the heat loss is calculated essentially for each square foot of the building's exposed surface area for a maximum design-heating hour. This would include windows, doors, skylights, foundation losses, and all outside exposed surfaces. Also taken into account are the heat load and gain from internal equipment, solar radiation/time-of-day incidence, and surface barrier conditions. Detailed descriptions of these location-sensitive calculations can be found in most Heating, Ventilation and Air Conditioning (HVAC) manuals. Using the design heating load determines the sizing of the boiler (or other device) required. The equipment efficiency losses must also be taken into account to meet the net heating requirements.

A quick simplified example of this type of problem is as follows:

- A one-story residential structure located in Central Islip, New York, is designed with wall *R*-values of 12. The sloped roof is not insulated, but the internal ceiling is insulated to an *R*-12 value. The building is built on a slab foundation. Assume that the slab, doors, and windows have a heat-loss total = 7000 BTUs/h for this example. The dimensions of the building are 30 ft by 50 ft with 8-ft-high ceilings. Using a 99.6% heating design DB temperature of 11°F and a boiler efficiency of 80%, how many BTUs of fuel per hour (Q) are required to feed this boiler at the design temperature?
- 1st step: $(70°\text{F} - 11°\text{F}) = 59°\text{F}$
- 2nd step: $R = 12\ (\text{h} \times \text{ft}^2 \times °\text{F})/\text{BTUs}$, so $U = 1/R$; $1/12 = 0.08333\ \text{BTUs}/(\text{h} \times \text{ft}^2 \times °\text{F})$
- 3rd step: surface area (A) = $(30' \times 2 \times 8') + (50' \times 2' \times 8') + (30' \times 50') = 2780\ \text{ft}^2$

Solid–Sus

- 4th step: $Q = UA\Delta t$; 0.08333 BTU's/(h×ft^2×°F)× 2,780 ft^2×59°F = 13,668 BTUs/h boiler output + 7000 BTUs/h (slab/door/windows) = 20,668 BTUs/h.
- 5th step: accounting for equipment efficiency; Solution = (20,668 BTUs/h boiler output)/(80% boiler efficiency) = 25,835 BTUs/h boiler input.

This result uses a standard boiler efficiency of 80%. Using a high-efficiency condensing-type boiler would raise that efficiency to about 90%, reducing fuel consumption. Of course, the current cost of a residential-sized condensing boiler is about double that of a cast iron-type boiler. This is a very straightforward problem; in larger buildings with more complicated duct systems, many other considerations would need to be accounted for.

Estimating annual fuel consumption can be calculated using Heating Degree Day methodology.[16] This method assumes that solar and internal gains for a structure will offset heat loss when the daily mean temperature is 65°F. The formula for this is

$F = (24(\mathrm{DD})q\mathrm{Cd})/(n(\mathrm{ti} - \mathrm{to})H)$,
where F = Quantity of fuel required, units depend on H
DD = Degree Days for the period desired (from local weather stations)
q = Design heat loss (BTUs—Step 5 of the previous problem)
Cd = Interim correction factor[15]
n = System efficiency
H = Heat value of the fuel

Example: using the above solved problem and the 5-year local average of 6000 HDDs, how many gallons per year of No. 2 oil will be consumed at a heating value of 130,000 BTUs/gallon?

- Step 1: Cd = 0.6, q = 20,668 BTUs/h, n = 80%, H = 130,000 BTUs/gallon, (ti − to) = 59°F
- Step 2: F = (24×6000 HDDs×20,668 BTUs/h×0.6)/ (0.8×59°F×130,000 BTUs/gallon)
- F = approximately 291 gallons of No. 2 oil per year

This result does not account for domestic hot water production or occupancy anomalies. These simplified problems are for demonstration purposes only.

CURRENT TRENDS

Recent historical events have changed the energy market dramatically. Fuel, once cheap and plentiful, has become limited in supply, causing costs to rise dramatically. The effect on lifestyles and economies has been drastic. The effect on technology has been an upswing in energy-efficient technologies and practices. Automobiles with high-efficiency engines and hybrid technologies have replaced powerful but inefficient V8 engines. Lighting has seen more use of fluorescent fixtures, which produce more lumens per watt than the older incandescent bulbs. More efficient condensing-type boilers have seen more use in the marketplace. Condensing boilers capture more heat out of the exhaust, resulting in lower exhaust temperatures and higher overall efficiencies. Many building codes have been modified to reflect the need for "green buildings." These "sustainable" or "high-performance" structures have a reduced impact on Earth's resources. They are designed to be comfortable yet resource efficient, and they will conserve energy, water, and raw materials. They minimize waste both in design and operation, and the life cycle of the building is also taken into account.[17]

Space heating improvements have been evident in the field of solar heating. These systems can be passive, active, or a combination of the two. Passive solar heating systems should take advantage of the design features of a habitat. If used, homes are built of materials such as tile and concrete. These materials absorb heat during the day and release this stored energy in the evening when it is needed. An active system consists of a collection device that absorbs the solar radiation, and in conjunction with fans and pumps, it transfers and distributes this heat. Some active systems may utilize an energy-storage system to provide heat when the sun has gone down. The two basic systems use either liquid or air to store this heat. The liquid systems heat either water or an antifreeze solution in a hydronic collector. Air-based systems use an air-to-water heat exchanger, dissipating the heat into the living space. During the summer months, the system is used to heat water for domestic use. Many of these systems can provide 30%–70% of the heating requirements of most homes.[17]

Another geographically limited means of space heating is geothermal energy. Iceland has a high concentration of volcanic activity and geothermal energy, which is put into wide use for heating and the production of electricity. The energy is so inexpensive that the sidewalks in Reykjavic and Akureyi are heated in the wintertime. About 17% of the country's electricity (219 MW) and 87% of the nation's heating and hot water requirements are supplied by this means.[18]

Improvements in window technology have improved dramatically over the past 100 years. The old single-pane glass windows were literally thermal holes in buildings. Their R-values ranged around 1.0, meaning that they allowed about 19 times more heat to escape per square foot than the average wall. With new insulated windows, there is a gaseous or vacuum barrier between the two sheets of glass, and the frames have been improved to essentially eliminate air infiltration. R-values in the range of 4.0 and higher have been met with current construction practices. This is a vast improvement over the older single- and double-pane storm-type windows. The insulating value of a double- or triple-pane window is a product of the air space between the panes and the type of spacers used.

Metal has been avoided for use as spacers, as it conducts heat and cold very well. Rubber, foam, and fiberglass have replaced metal in most applications.

The frames and materials have progressed where many architectural styles and many materials are available. Some of the materials are aluminum, wood, aluminum clad, vinyl clad, wood–plastic composites, vinyl, and fiberglass. Since the Energy Policy Act of 1992, the National Fenestration Rating Council (NFRC) has worked on a national window-standard testing procedure. It has also set the standard for rating windows according to solar heat gain, air infiltration, and condensation potential.[19]

Originally, most HVAC controllers were pneumatic. Mechanical engineers used their experience with steam and air to control hot or cold air. In time, with standardization, relay ladder logic began to replace pneumatic controls, which in turn were replaced by electronic switches. Modern computerized control systems can now be accessed by Web browsers with the appropriate safety passwords.[20] These controls can remotely control temperature, humidity, and pressure. Many of these systems also include lighting, fire, and other security controls. The use of programmable time-of-day controls allows any building to maintain comfortable set points in accordance with ASHRAE standards for day-time-occupied and evening and weekend operation. In most applications, the building temperatures will be allowed to drop during nonoccupied hours, generating fuel savings. This set point must take into account ASHRAE standards and individual building requirements. Shortly before the building opens, heat will be supplied to bring the comfort level up to normal. At the end of the day, the temperature will be allowed to drop back slowly to the evening set point. Residential use of this technology works much the same way. Occupancy sensors have also been used for lighting and to maintain lower set points in larger unoccupied areas.

CONCLUSION

With the current world energy situation, energy conservation and high efficiency are now household words. Technological advances are being investigated to increase efficiencies and cost effectiveness in many fields of study. This is being attempted while maintaining the lifestyles that we are accustomed to. The public need to be more aware of how it uses energy to reduce dependency on foreign fuels. It is usually during times of adversity that some of the more intuitive insights are realized and great achievements are accomplished.

In the middle of difficulty lies opportunity

-Albert Einstein

REFERENCES

1. http://www.romans-in-britain.org.uk/inv_central_heating.htm (accessed May 2006).
2. http://www.vroma.org/~bmcmanus/baths.html (accessed May 2006).
3. http://www.romans-in-britain.org.uk/inv_central_heating.htm (accessed May 2006).
4. http://www.castlewales.com/life.html (accessed April 2006).
5. http://en.wikipedia.org/wiki/Kitchen (accessed April 2006).
6. http://www.vtbrickoven.com/masonry/masonry.html (accessed June 2006).
7. http://www.secretlifeofmachines.com/secret_life_of_central_heating.shtml (access May 2006).
8. http://www.egr.msu.edu/~lira/supp/steam (accessed May 2006).
9. http://www.egr.msu.edu/~lira/supp/steam (accessed May 2006).
10. http://www.egr.msu.edu/~lira/supp/steam (accessed May 2006).
11. http://www.hsb.comabout.asp?id-50 (accessed April 2006).
12. http://www.memagazine.org/backissues/february2000/features/setting/origin.html (accessed May 2006).
13. http://www.elizabethtowngas.com/BusinessCustomers.CommercialBoilers.aspx (accessed April 2006).
14. ASHRAE Climatic design information. In *1989 ASHRAE Handbook—Fundamentals—I-P Edition*; American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc.: Atlanta, GA, 1989; 26-6–26-21 [Ch 26-6].
15. ASHRAE. *In 1989 ASHRAE Handbook—Fundamentals—I-P Edition*; American Society of Heating , Refrigeration and Air-Conditioning Engineers, Inc.: Atlanta, GA, 1989.
16. http://www.moea.state.mn.us/greenbuilding/index.cfm (accessed May 2006).
17. http://www1.eere.energy.gov/solar/sh_basics_space.html (accessed May 2006).
18. http://www.en.wikipedia.org/wiki/Geothermal_power_in_Iceland (accessed June 2006).
19. http://www.healthgoods.com/EducationHealthy_Home_Information/Doors_and_Windows (accessed May 2006).
20. http://en.wikipedia.org/wiki/HVAC_control_system (accessed May 2006).

Solid–Sus

Steam and Hot Water System Optimization: Case Study☆

Nevena H. Iordanova
Senior Utility Systems Engineer, Armstrong Service, Inc., Orlando, Florida, U.S.A.
Rick Cumbo
Operations and Maintenance Manager, Armstrong Service, Inc., Orlando, Florida, U.S.A.

Abstract
In 2001, Armstrong Service performed an audit in a large food processing plant.
The audit focused on the steam and hot water systems improvements, as well as their monitoring, control and sustainability of improvement results.
The objective of this engineered audit was to:

- Improve the reliability of the existing systems, including generation, distribution, use, and condensate return,
- Improve system performance to achieve better efficiency,
- Reduce the steam usage, venting, condensate drain and heat loss from the system,
- Reduce emissions and effluents to the environment,
- Improve monitoring and measurement of the system parameters and assure long lasting results.

The presentation will highlight:

- The engineering analysis of the existing system. The boiler house operates with three natural gas fired boilers consuming up to 310,000 MMBTU/year. Saturated steam is generated at 115 psig and distributed to the plant at an average flow rate between 20,000 and 50,000 lbs/h to multiple process users.
- Identifying and root cause analysis of the deficiencies/problems that cause energy losses,
- Providing the necessary remedial actions to improve the steam and condensate system to reduce energy use and maintenance cost. The overall assessment and analysis resulted in 19 conceptual energy conservation measures (ECMs), with potential to reduce up to 30% of the steam cost. After being presented to the plant, 9 out of 19 proposals were approved for further development. In 2002, the projects were implemented and the plant steam to product efficiency was improved by more than 10%.
- Monitoring/managing the system and sustaining the results.

Additional to the instrumentation and meters installed to collect data on critical points for each individual project, a web enabled monitoring and reporting energy optimization system (EOS) was installed. It made the process of data acquisition, monitoring, recording, trending, reporting, and control possible not only for the plant operation and maintenance personnel, but also accessible from a standard web browser for a control and management prospective.

INTRODUCTION

Improving steam system efficiency will contribute significantly to the profitability for every plant. Nearly half of industrial energy is used to generate steam. There are proven energy savings techniques that capture significant energy saving benefits upon implementation. Some of them are industry wide; others are specific for each plant.

Steam systems consist of several components such as:

- Steam generation
- Steam distribution
- Steam usage
- Condensate collection and return

A comprehensive steam system optimization addresses these interrelated areas, identifies the problems, and recommends all necessary corrective measures to eliminate them.

OVERVIEW OF THE STEAM SYSTEM

Steam Generation Cost

The annual steam generation cost at the Plant is approximately $1,766,000, consuming 310,000 MMBTU/

Keywords: Steam systems; Steam system optimization; Food processing heat; Manufacturing steam use; Steam traps; Condensate collection; Waste heat recovery; Boiler blowdown; Steam cookers.

Table 1 Steam generation cost

Utilities		Unit costs		Incremental steam cost	
	Units/yr	**Units**	**$/unit**	**$/yr**	**$/klb steam**
Fuel	309,860	MMBTU	5.24	1,623,666	5.86
Water	21,600	kgal	1.96	42,349	0.15
Sewer	4,324	kgal	4.04	17,469	0.06
Chemicals				31,092	0.11
Electricity	1,289	MWh	40	51,564	0.19
Average steam cost				1,766,139	6.37

year. This cost represents the total cost that is required to generate an average of 25,000 lbs/h of steam at 115 psig. The various costs that make-up the steam generation costs are shown in Table 1.

Steam Generation System

There are three boilers in service in the facility:

- Murray built in 1960 rated at 60,000 #/h
- Union built in 1966 rated at 125,000 #/h
- Murray built in 1972 rated at 100,000 #/h.

All of the boilers use natural gas as their primary fuel. Typically, one or two of the boilers are running with one other boiler in standby (Fig. 1).

Steam Utilization

The main steam users are the product cookers, space heaters and the steam jet ejectors on the product deaerators (DAs). An extremely large amount of heat is being wasted by the venting of steam from the vacuum jet ejectors (see ECM#3). Steam is also used to a lesser extent for sterilization, in the jacketed reheat vessels and for water heating.

Steam Trapping

Results from the steam trap survey performed during the audit are listed in ECM #5. In addition to the defective traps, there are several other notable problems with steam traps at the facility. A large amount of steam is being wasted in order to drain condensate from some of the product equipment (see ECM #6).

Condensate Return System

Condensate is piped back to a main condensate receiver (CR) in the production area and then pumped back to the

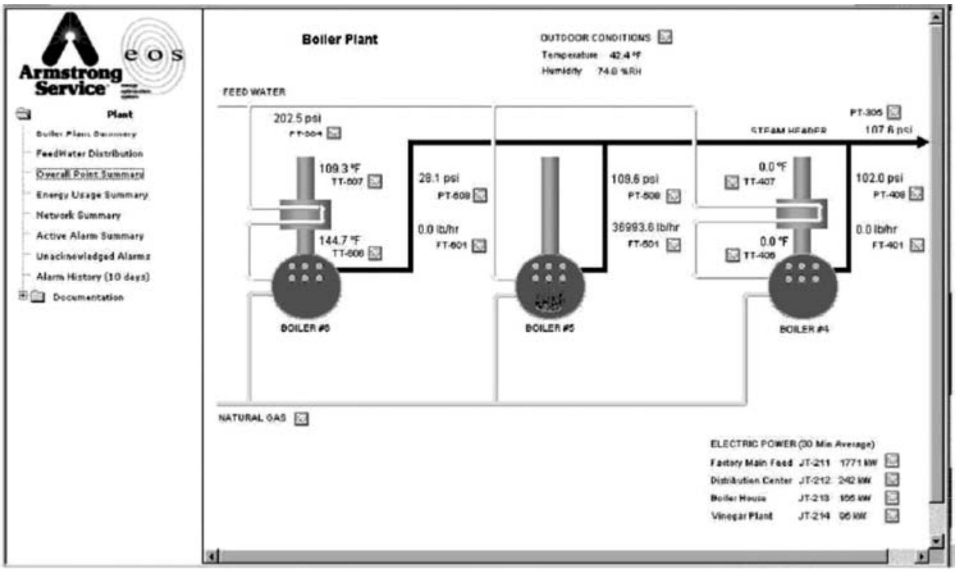


Fig. 1 Steam generation system.

boiler house. Flash steam off this CR is piped back to the boiler house and used to heat makeup water to the DA.

Waste Heat Recovery

Some of the waste heat is recovered by the use of the flash steam in the DA and the use of economizers on several boilers.

SAVINGS OPPORTUNITIES

A thorough review of the system confirmed that there are energy saving potentials in all the areas of the steam system: steam generation, distribution system, utilization, and in condensate return system. The following paragraphs highlight nine energy savings opportunities identified at the site and describe the measures proposed and implemented in order to realize these savings.

ECM 1: INSTALLATION OF A STACK ECONOMIZER

Current System Description

The boiler with the highest design capacity, 125,000 lbs/h, is currently being used as an alternate base load boiler, averaging about 40% of the operating time per year. This boiler is usually used during periods of expected higher steam demand. The boiler does not have an economizer and the flue gas temperature is around 520°F.

ECM Description

Armstrong Service, Inc. (ASI) proposed to install one fin tube stack economizer, which would preheat the boiler feedwater with the heat recovered from the flue gas (Fig. 2). The installation of the economizer will improve the boiler efficiency by reducing the exhaust flue gas temperature down to 285°F. After the installation of the economizer, the boiler will be the primary boiler for the plant and will be used 90% of the time.

The boiler is normally fired on natural gas. However, it has the capability of being fired on Fuel Oil #6. Therefore, an automatically operated soot blower will be installed to help maintain the economizer.

Proposed Energy and Utility Savings

In FY 2002, the steam generation at the plant was 277,151 Mlb. The energy savings expected from the economizer installation were based on 94% of the present steam load, as the load was expected to be reduced down to approximately 260,000 Mlb after the repair of the major non-design vents and leaks.

More details about the data used and the saving calculations are included in Table 2.

By reducing the fuel consumption, total carbon emissions were reduced by 100 tons/year.

The simple payback for this project was 3.5 years.

ECM 2: REPLACEMENT OF BLOWDOWN HEAT RECOVERY EQUIPMENT

Current System Description

All boilers are operated with a continuous blowdown. Each unit has a manual valve, which is used for adjustment. A blowdown heat recovery unit was previously installed in the Boiler House in the DA penthouse. It consisted of a flash tank and a heat exchanger. In the flash tank the blowdown flashes to a lower pressure equal to the DA pressure and supplements the low pressure (LP) steam required at the DA. The heat exchanger recovers the heat from the hot liquid by preheating boiler make-up water (MUW).

The blowdown heat recovery unit has been disconnected and abandoned in place, due to persistently occurring leaks at the heat exchanger and the excess of LP steam to the DA from the plant CR vent line.

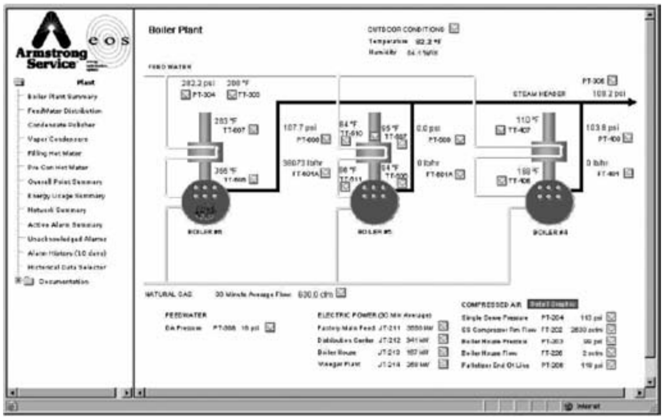


Fig. 2 Proposed economizer.

Table 2 Savings from economizer installation

Economizer savings calculation		Existing condition		Proposed modification	
Hours of operation: 24 h/day 6 days/ week 50 weeks/year 7200 h/year		B#5	40%	B#5	90%
		B#6	60%	B#6	10%
		B#5	B#6	B#5	B#6
		Existing	Existing	New	Existing
		No Econ	W/Econ	W/Econ	W/Econ
Input	Unit				
Fuel selection	Type of fuel	NG	NG	NG	NG
Fuel cost	$/MMBTU	5.24	5.24	5.24	5.24
Maximum design firing rate	MMBTU/h	125	80	125	80
Steam load	lb/h	36,001	36,001	36,001	36,001
Steam pressure	psig	115	115	115	115
Feedwater temperature	°F	220	220	220	220
Ambient temperature	°F	80	80	80	80
Flue gas temperature	°F	520	350	285	350
Intake air temperature	°F	100	100		
Flue gas analysis based on measurement (% vol—dry basis)					
	O2	3.5	3	3.5	3
	CO	0	0	0	0
	Combustibles	0	0	0	0
Annual operating hours	h	2,880	4,320	6,480	720
Heating value	Btu/ft^3	1,001.707	1,002	1,001.71	1,002

		Before	After
Summary of Results			
Fuel consumption	MMBTU/yr	314,706	308,020
Fuel cost	$/yr	$1,649,062	$1,614,027
Carbon emissions	ton/yr	4,716	4,615
Savings			
Fuel consumption	MMBTU/yr	6,686	
Fuel cost	$/yr	$35,035	
Carbon emissions	ton/yr	100	

ECM Description

ASI proposed to replace the blowdown heat recovery system. The waste heat in the blowdown water was used to preheat boiler makeup water (Fig. 3). The flash tank was not reused and the blowdown water was recovered at a high pressure in a shell and tube system of counter flow heat exchangers. The blowdown discharge temperature would be 170°F, due to the MUW temperature coming in at 140°F.

Proposed Energy and Utility Savings

Based on the conductivity readings from the feed water and the boiler water, the average blowdown rate was calculated on 9.4%.

More details about the data used and the saving calculations are included in Table 3.

By reducing the fuel consumption, total carbon emissions will be reduced by an estimated 70 tons/year.

The simple payback for this project was 1.5 years.

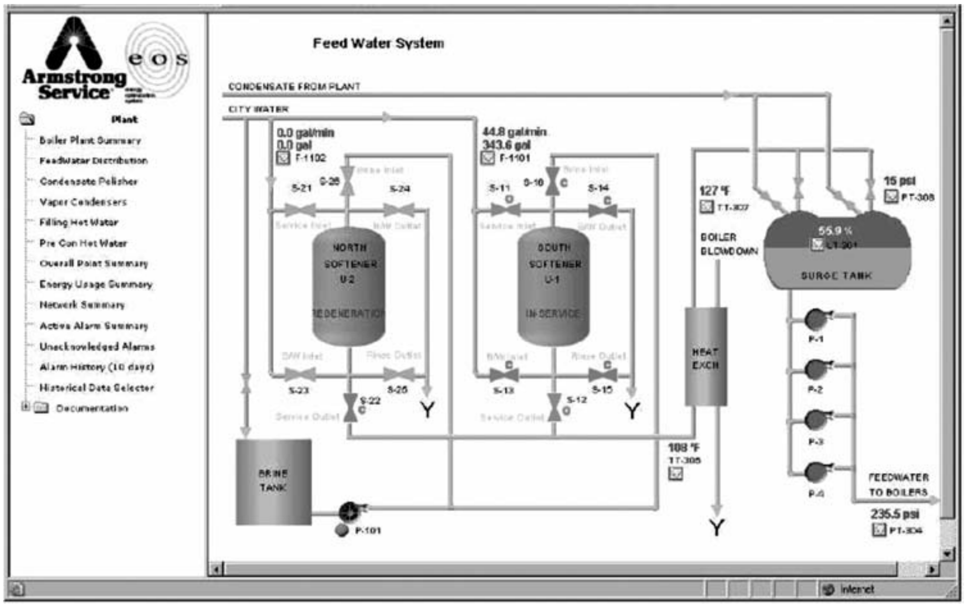


Fig. 3 Blowdown heat recovery.

ECM 3: REPLACEMENT OF PRODUCT DEAERATOR STEAM JET EJECTORS WITH MECHANICAL VACUUM PUMPS

Current System Description

Deaeration is one of the steps of the food product processing. The product is sprayed in a DA at 205°F at a rate of 60 gpm. Due to the vacuum, part of the water is evaporated and part of the oxygen is removed. The product then leaves the DA at 190°F through a process pump.

The plant has a total of eight (8) existing product DAs. The vacuum is drawn through barometric condensers by steam jet ejectors (exhausters) designed to maintain up to 15′ Hg vacuum. Each steam ejector is rated for a dry air suction capacity of up to 70 lb/h. The arrangement of the existing system is shown in Fig. 4.

Table 3 Blowdown heat recovery saving

Annual steam generation	259,204	Mlb/yr
Annual operating hours	7200	hr/yr
Average steam generation	36,001	lb/h average
Blowdown	9%	
Blowdown quantity	3240	lb/h
Pressure	115	psig
Blowdown heat	318.7621	Btu/lb
Blowdown temperature to drain	180	°F
Heat recovered	0.55	MMBTU/h
Steam generation efficiency	85%	
Fuel savings	4,687	MMBTU/yr
Sewage increase	−401	Mgal/yr
Heat savings steam equivalent	4232	klb
Annual dollar savings	$22,937	$/yr

ECM Description

ASI proposed to replace some of the steam jet ejectors with electrically driven liquid ring (LR) vacuum pumps (VP) and the direct water-cooled condensers with indirect

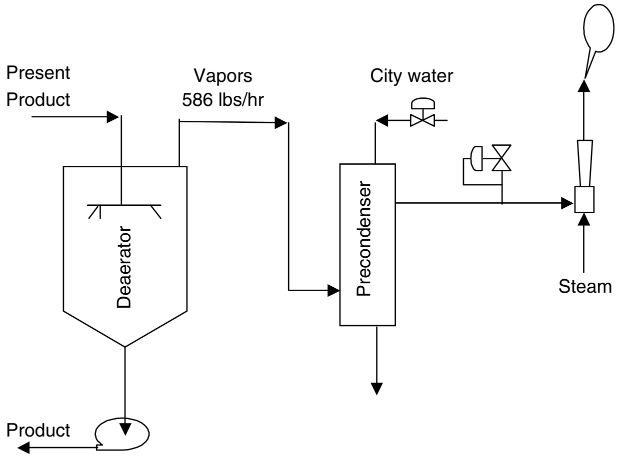


Fig. 4 Existing equipment configuration.

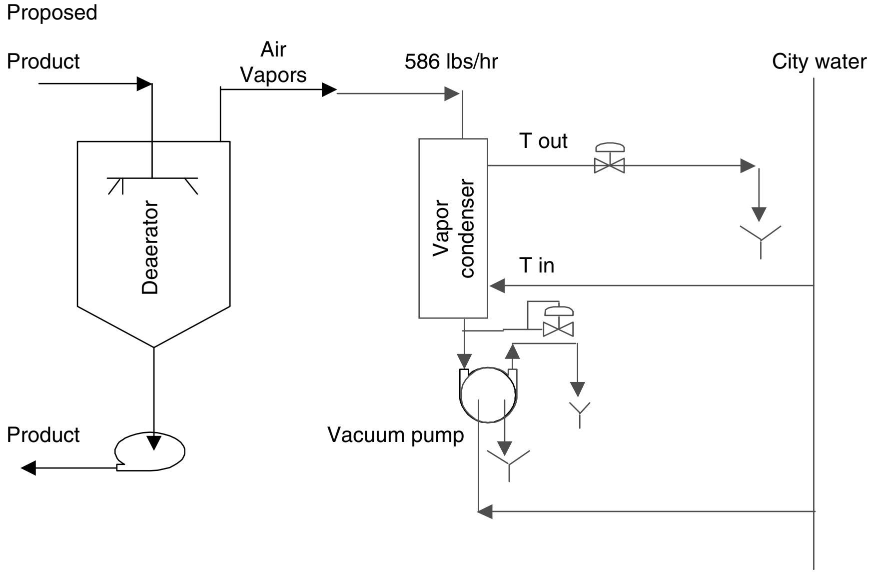


Fig. 5 Proposed equipment configuration.

condensers at three of the DAs. The VPs will eliminate the need for steam use in the DAs and the indirect condensers will eliminate the water discharge to industrial waste sewer (IWS). A simplified schematic of the proposed piping diagram is shown in Fig. 5.

Energy and Utility Savings and Environmental Impact

Under the current conditions each steam ejector was using 357 lb/h of steam and each condenser discharged approximately 15 gpm of cooling water to the IWS.

Details about the data used and the savings calculations for one of the DA are included in Table 4. Table 5 shows the summary of the savings from all the reconfigured DAs.

By reducing the steam usage, total carbon emissions were reduced by 100 tons/year.

The installation of the new design for the product DAs needed to be completed around the work schedule of the current product DAs. The simple payback for this project was 4 years.

ECM 4: REPAIR AND/OR REPLACE RELIEF VALVES AND CONTROL VALVES

Current System Description

There are pressure relief valves and atmospheric control/relief valves located downstream of the control valves on the steam inlet line for multiple cooking kettles. These relief valves are connected to common vents, which are then piped through the wall to the outdoors. Steam is being discharged from these vents approximately 60%–70% of the time (Fig. 6).

An ultrasonic check of the control valves proves that six of the atmospheric valves, all of the pressure relief valves and seven of the steam supply valves are leaking steam.

ECM Description

ASI proposed to replace and/or repair the leaking valves.

Energy and Utility Savings and Environmental Impacts

The unnecessary steam venting was driving up the cost of steam. The boiler fuel usage increased, as more fuel had to be used to supply the additional steam load. The steam lost to the atmosphere increases the amount of MUW required, as it is not recovered as condensate. The additional MUW requires preheating and water treatment/chemicals. The energy savings for this ECM result in 5,223,000 lbs of steam per year and the details can be seen in Table 6.

By reducing the steam venting, the total carbon emissions were reduced by 189 tons/year.

The simple payback for this project was 2.8 years.

ECM 5: STEAM TRAP REPAIR AND REPLACEMENT

Current System Description

ASI conducted a steam trap survey at the plant and 180 traps were tested. The survey and computer analysis revealed that approximately 25% of the in-service traps were defective. The defective traps were wasting steam

Table 4 Utility consumption comparison

Present operation		
Deaerators		Present
Product flow	gpm	XX
Product temperature In	°F	205°F
Product temperature Out	°F	190°F
Vacuum maintained in DA	In HG	10–15
Precondensers		
Water wasted in condensers	gpm	14.8
City water temperature	°F	65
Hot well temperature	°F	130
Steam jets (Exhausters)	psig	100
Steam used	lb/h	357
Utilities required to operate DA present		
City water+MUW (city)	gpm	15.6
Sewer	gpm	14.8
Steam	lb/h	357
Fuel	Btu/h	474453
Proposed operation		
Vacuum liquid ring pumps		
El pump power required	HP	5
Pump water seal flow	gpm	1
Precondenser		
City water In	°F	65
City water Out	°F	105–120
Calculated condenser heat load	Btu/h	500,000
Savings from one DA		
Fuel		
MMBTU/yr	1358	
City water (all=cooling, seal, MUW)	Mgal/yr	−1796
Industrial sewer	Mgal/yr	2195
Chemicals	$/yr	$143
Electricity	kWh/yr	−10675

Table 5 Utility savings—ECM 3

Type	Units	Units/yr	$/yr
Fuel	MMBTU	5,333	27,947
City water	1000 gal	−7,053	−13,828
Waste water	1000 gal	8,620	34,825
Chemicals	Lot	1	562
Electricity	KWh	−41,930	−1,677
Total	$		47,828

Fig. 6 Leaking relief valves vent.

in excess of 2000 lb/h. In addition to the traps wasting steam, 3.7% of these traps were failed in a cold plugged and/or flooded conditions. These traps are not losing steam but rather causing a loss of heat, water hammer, corrosion, and possibly damage to heat transfer equipment. The following is a break down of defective traps as a percentage of the total in-service traps:

Blow-Thru	21.5%
Cold plugged	3.7%
Total	25.2%

ECM Description

The existing steam traps were very old and many of them need to be replaced. Installing the proper replacement trap will improve the systems efficiency. ASI proposed to replace/repair/modify any failed traps or incorrect piping configurations.

Table 6 Utility savings—ECM 4

Type	Units	Units/yr	$/yr
Fuel	MMBTU	7112	37,266
City water	1000 gal	627	1,229
Chemicals	Lot	1	668
Total	$		39,163

Table 7 Utility savings—ECM 5

Type	Units	Units/yr	$/yr
Fuel	MMBTU	4,028	21,107
City Water	1,000 gal	362	709
Chemicals	Lot	1	385
Total	$		22,201

Energy and Utility Savings and Environmental Impacts

By replacing the defective steam traps, 3,012,000 lbs of steam were saved. The details about the savings can be seen in Table 7.

By reducing the steam venting, total carbon emissions were reduced by 60 tons/year.

The simple payback for this project was 1.8 years.

ECM 6: CORRECT STEAM BLOWTHROUGH RATE ON COOKER KETTLES

Current System Description

The cooking kettles are steam jacketed type vessels or have steam heating coils. The steam heating coils in nine of the cooking kettles use float traps (liquid drainers). Steam is supplied to two helical copper coils in the bottom of the vessel. In order to heat up the coils as quickly as possible, a bypass valve ahead of the each trap is open to the condensate return line. During the cooking phase a large amount of steam flow passes directly through this valve and into the return line.

ECM Description

ASI proposed to replace the existing liquid drainers with steam traps that had been designed with a fixed restricted orifice near the top of the body, built into it to bleed off the steam and non-condensable. This would allow the steam bleed rates to be adjusted to achieve maximum performance with minimal steam blow flow.

Table 8 Utility savings—ECM 6

Type	Units	Units/yr	$/yr
Fuel	MMBTU	6,343	33,239
City water	1,000 gal	569	1,115
Chemicals	Lot	1	607
Total	$		34,961

Energy and Utility Savings and Environmental Impacts

By installing the appropriate steam traps for this application, 4,743,000 lbs of steam were saved. The savings were calculated based on 28 psig steam pressure and 60,000 h of operation. The details about the savings can be seen in Table 8.

By reducing the steam venting, the total carbon emissions were reduced by an estimated 95 tons/year.

The simple payback for this project was 4.1 years.

ECM 7: PRODUCT COOLING HEAT RECOVERY AND WATER RE-USE

Current System Description

The plant has multiple product lines equipped with coolers (plate heat exchangers) to precool the product from 190 to 85°F before the fillers. City and/or well water is used as the cooling media. After the coolers, the cooling water (ranging from 145 to 170°F) is discharged to a clear water sewer (CWS) sump. Every shift, samples of the cooling water are taken and tested for the purity and pH of the water. If the samples are contaminated with product, the water is diverted to a common header and then discharged to the IWS.

At the same time, cold city water is used as an ingredient in all feedstock preparation and cooking processes. For many of the water users, heating and/or cooking takes place after the mixing of the ingredients. Steam at 115 psig is used as the heating media.

In the boiler house, cold city water is used as boiler MUW.

ECM Description

ASI proposed to use the hot water discharge from the existing product coolers (155°F average used for savings) for make up-water to the product users. The remainder of the hot water discharge would be used as boiler MUW.

The newly proposed system was designed in details (Fig. 7). After the plant approval, the ECM was implemented and the performance of the equipment and the savings are monitored through the energy optimization system (EOS) system since then (Fig. 8).

Hot water discharge, from the coolers, which use only city water, was collected into two separate tanks. From there, the hot water is pumped to the different users as needed. The pumps are equipped with variable frequency drives (VFD) to efficiently handle the large variations of the flow rates. Each pump has equivalent capacity back-up

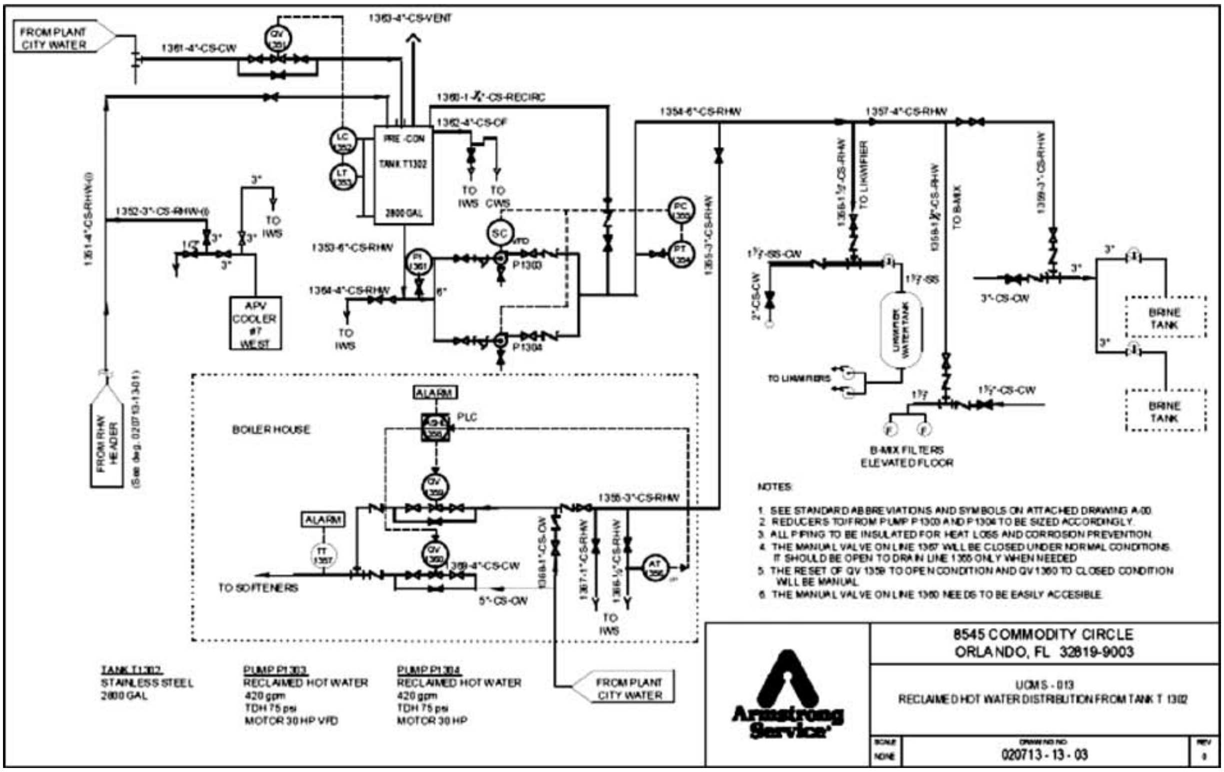


Fig. 7 ECM 7 sample P&ID—TANK 1.

to assure process reliability. Both tanks will have a city water back-up system to assure water availability at any time.

In case of interruption of the Boiler MUW supply, the transition to city water supply, as MUW, is automatic.

Energy and Utility Savings and Environmental Impacts

The implementation of this ECM saved 36,212,000 lbs of steam, shortened the process heat-up time and increased the yield rate without any change of the process equipment.

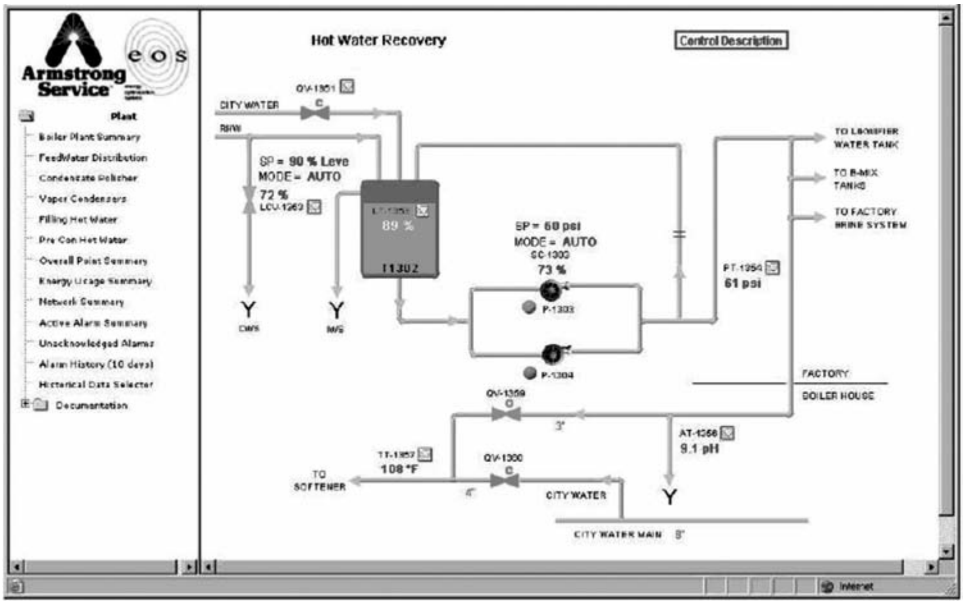


Fig. 8 EOS monitoring screen—ECM 7.

Table 9 Utility savings—ECM 7

Type	Units	Units/yr	$/yr
Fuel	MMBTU	37,857	198,371
City water	1,000 gal	40,525	79,455
Waste water	1,000 gal	−600	−2,424
Electricity	KWh	−232,128	−9,285
Total	$		266,117

Table 10 Utility savings—ECM 8

Type	Units	Units/yr	$/yr
Fuel	MMBTU	1,964	10,293
City water	1,000 gal	173	340
Chemicals	Lot	1	185
Total	$		10,817

The savings were calculated based on city water temperature 65°F, average for the year, heated up to 155°F. The details about the savings can be seen in Table 9.

By reducing the steam usage, the total carbon emissions were reduced by 567 tons/year.

The simple payback for this project was 2.4 years.

ECM 8: ISOLATE ABANDONED LINES

Current System Description

The leak through a strainer ahead of an abandoned valve has been wasting steam for 14 years. Steam is currently being unnecessarily vented through a 2′ pipe, next to an entrance door.

ECM Description

ASI proposes to isolate the leaky strainer and valve.

Energy and Utility Savings and Environmental Impacts

The implementation of this ECM saved 1,443,000 lbs of steam. The details about the savings can be seen in Table 10.

By reducing the steam venting, total carbon emissions will be reduced by an estimated 29 tons/year.

The simple payback for this project was 0.1 years.

ECM 9: CONDENSATE RETURN IMPROVEMENT

Current System Description

The Condensate Return System is a closed system. The condensate from the plant is collected in a common CR. The CR has a pressure equalizing vapor line connecting it to the DA. Normal pressure maintained at the CR ranges from 10 to 20 psig, which is equal to the DA. The liquid condensate at the same pressure is pumped through a pump to the DA.

On average, 50% of the generated steam is returned as condensate (30–60 gpm). Due to the nature of the cooking process, occasionally the condensate gets contaminated and needs to be dumped. The contaminations are caused by hardness from cooling water

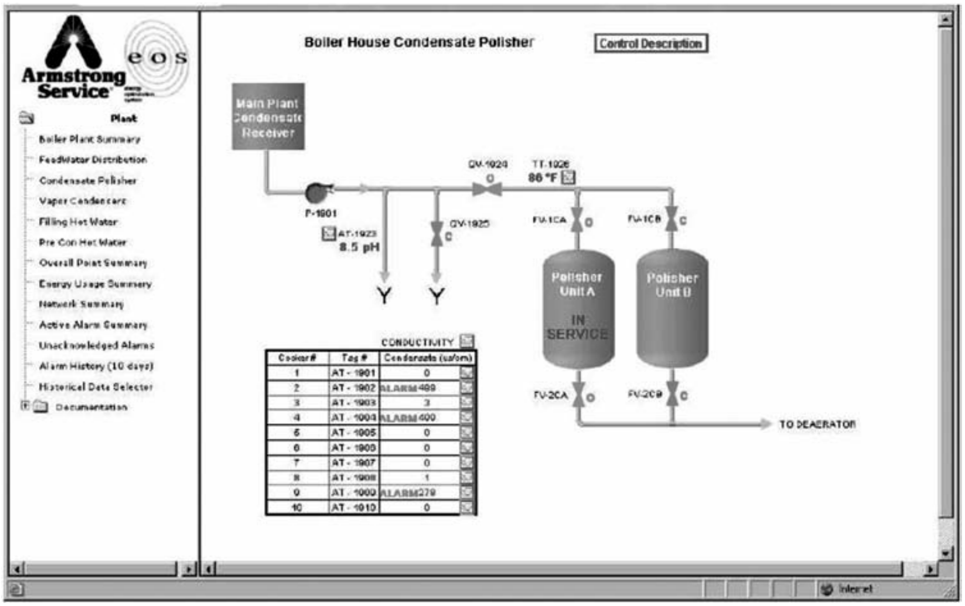


Fig. 9 EOS monitoring screen—ECM 9.

Table 11 Utility savings—ECM 9

Type	Units	Units/yr	$/yr
Fuel	MMBTU	10,699	56,064
City water	1,000 gal	5,326	10,435
Waste water	1,000 gal	5,006	20,212
Chemicals	Lot	1	5,681
Total	$		92,392

crossovers and/or by pH and conductivity from ruptured product coils. The condensate is drained at an estimated 1800 h/year and involves several hours of labor for each occurrence.

There was no arrangement for a fast and effective determination of the location of the contamination source. In addition, there was no way to separate the contaminated condensate from clean condensate. Due to this, all of the collected condensate was dumped. The condensate continued to be dumped until the source of contamination was found and repaired.

Table 12 Measurement and verification gross summary table

ECM	ECM title	Goals and objectives of M&V plan	Baseline measurements	Energy savings calculation	Verification measurement tool	Potential EOS monitoring points
1	Install boiler stack economizer	Calculate savings due to boiler stack exhaust/boiler feedwater pre-heat	Stack temperature log, hours of operation	Preheat feed water	EOS monitoring	Economizer flow, temperature in, temperature out
2	Blowdown heat recovery	Calculate savings due to reusing steam a 2nd time and preheating boiler feed water	Flow of condensate	Blowdown water heat recovery; Btu content of blowdown	EOS monitoring	Heat exchanger flow, temperature in, temperature out
3	Install vacuum pumps	Calculate savings due to steam reduction; increased electrical consumption	Hours of operation; steam flow; water flow; electrical energy	City water flow rate; electrical energy usage	EOS monitoring	Electrical metering; water flow; vapor condensor temperature in, temperature out
4	Relief and control valve replacement	Calculate savings due to steam reduction	Hours of operation; steam flow	Steam plume; condensate return	N/A	N/A
5	Steam trap replacement	Calculate savings due to steam trap replacement and system optimization	Test steam traps for leakage	N/A	Annual trap survey	N/A
6	Steam bleed on product cookers	Eliminate live steam from condensate return system	Boiler house deaerator pressure	Steam plume from DA	EOS monitoring	DA tank pressure

(Continued)

Solid–Sus

Table 12 Measurement and verification gross summary table *(Continued)*

ECM	ECM title	Goals and objectives of M&V plan	Baseline measurements	Energy savings calculation	Verification measurement tool	Potential EOS monitoring points
7	Hot water reuse—product cooling	Confirm reduction in city water use and heating due to pre-heated water re-use	Hours of operation; steam flow; city water flow	Reduced city water consumption, Btu content of hot water reclaimed	EOS monitoring	Heat exchanger flow, tank temperature, city water flow measurements
8	Isolate the abandoned lines	Calculate savings due to steam reduction	Steam flow	Steam plume	Visual	N/A
9	Condensate Return	Calculate savings due to steam reduction; volume of quality condensate return	Hours of cooker operation; condensate return flow; city water	Reduced city water consumption; Btu content of condensate	EOS monitoring	Polisher (conductivity/pH/flow/temperature)

ECM Description

ASI proposed to establish a monitoring system, which helps to identify the source of contamination and prevent continuous and larger than required dumping. ASI also proposed to install a Polisher, which assures good condensate quality before it is sent to the DA (Fig. 9).

A common pH probe automatically monitors and controls a pair of kettle cookers. An alarm is sent to the control room and an operator defines which one of the Cookers is leaking and manually closes the valves. By implementing this project, the draining of condensate was minimized to maximum two cookers for a short period of time.

A new condensate header was built, parallel to the Product cookers, to handle the contaminated condensate. The contaminated condensate was directed to a Flash Tank. The liquid was then discharged to the closest IWS location. A Condensate Polisher was installed in the boiler house to remove hardness from the condensate, which is

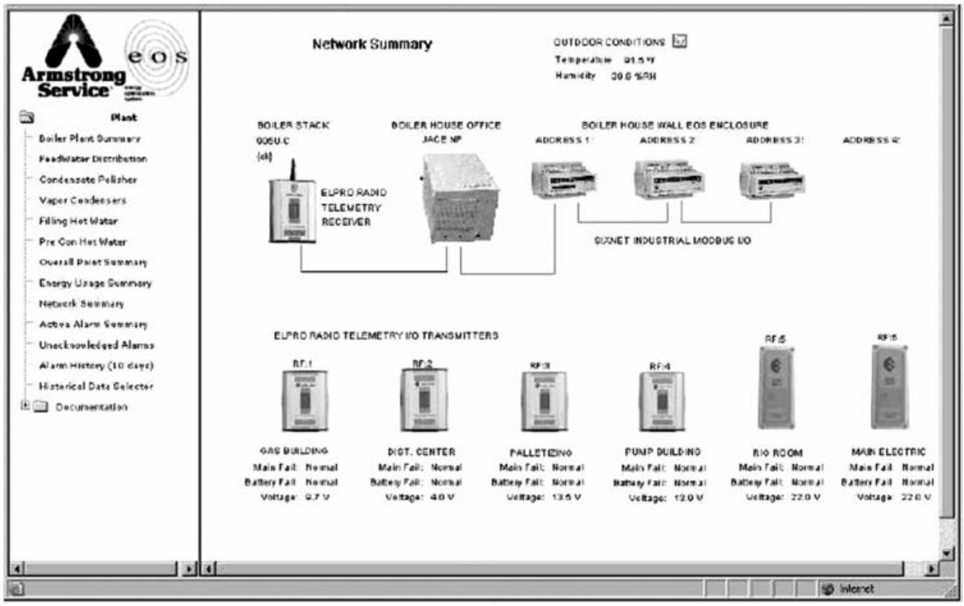


Fig. 10 EOS network summary.

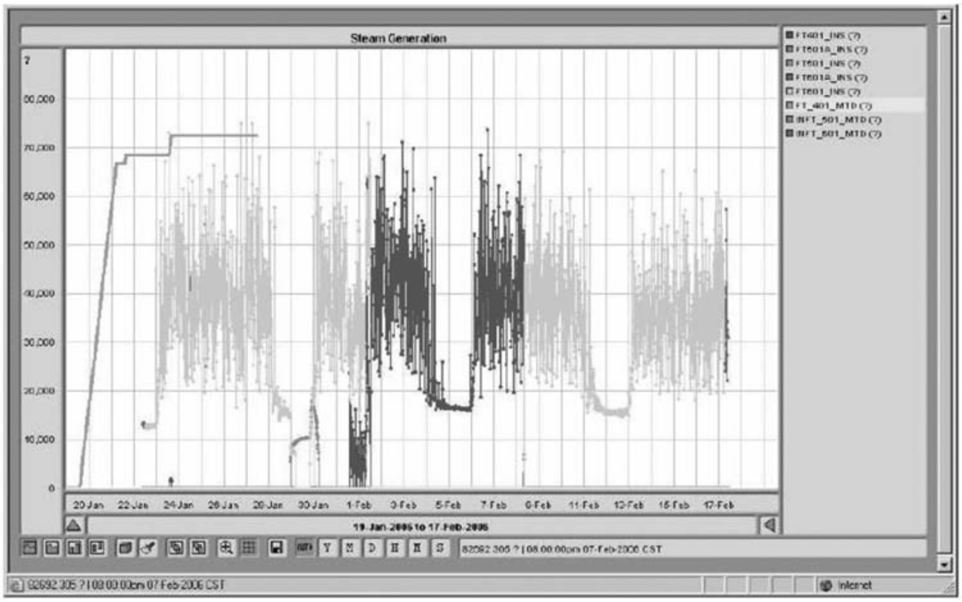


Fig. 11 EOS historic data selection.

coming from different cooling water crossover sources throughout the plant.

If the contaminations are extreme the condensate would be manually dumped to the IWS.

Energy and Utility Savings and Environmental Impacts

The implementation of this ECM saved 9,635,000 lbs of steam. The details about the savings can be seen in Table 11.

By returning the condensate, the total carbon emissions were reduced by an estimated 160 tons/year.

The simple payback for this project was 4 years.

MEASUREMENT AND VERIFICATION PROTOCOL

Before any of the above ECMs were implemented it was mutually agreed with the plant how the savings will be measured and/or calculated. Some of the savings were

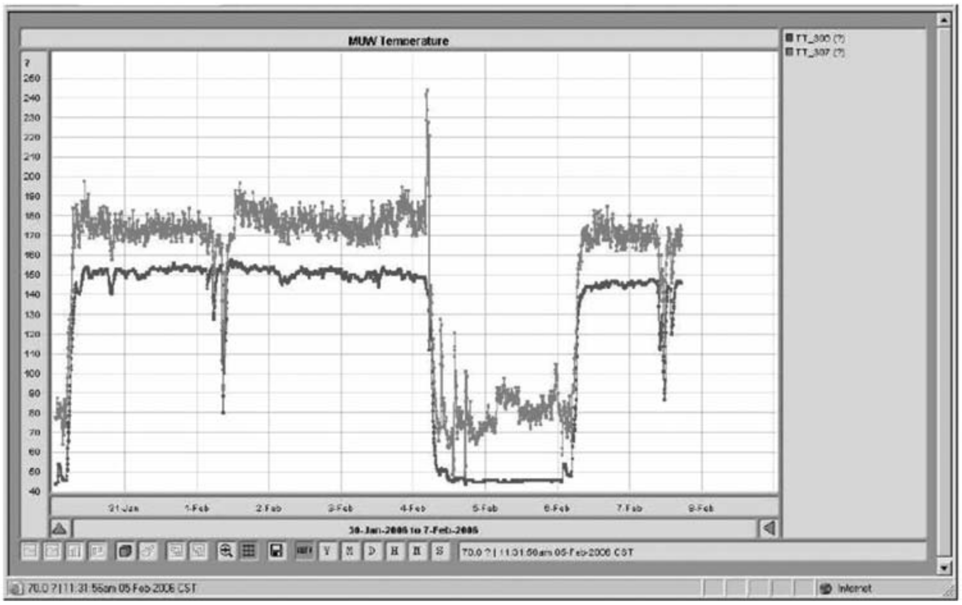


Fig. 12 EOS project performance evaluation.

Table 13 Summarized list of energy conservation measures

ECM	ECM title	Savings, $/yr	Payback/ yr	Carbon emissions, tons/yr
1	Installation of a stack economizer	35,000	3.5	100
2	Blowdown heat recovery	22,900	1.5	70
3	Install vacuum pumps	46,300	4.0	100
4	Repair relief and control valve	65,600	2.8	189
5	Steam trap repair/replacement	22,200	1.8	60
6	Steam bleed on cooker kettles	35,000	4.1	95
7	Product cooling hot water reuse	266,100	2.4	567
8	Isolate the abandoned lines	10,800	0.1	29
9	Condensate return	87,400	4.0	160
	Total savings/emissions	566,000		1,370

calculated based on the readings from the EOS and others were based on stipulated savings. Table 12 provides information on each proposal and the approach taken to establish a proper M&V protocol.

ENERGY OPTIMIZATION SYSTEM

An EOS was installed in the boiler house for monitoring/managing all the utility systems. The EOS was installed in addition to the instrumentation and meters installed to collect data on critical points for each individual project. Summary of the network is shown in Fig. 10.

Based on the existing or newly installed meters, with the EOS, historic data selection is easily manageable and presentable in a chart and/or a table format, depending on the user's preference. Fig. 11 shows the steam generation flow for the period of January 20th to February 17th, 2006.

By the installation of this web enabled monitoring and reporting system, it was possible to monitor the performance of the equipment, at any time and virtually from anywhere. Fig. 12 shows the performance of the heat recovery system (ECM 7) and the blowdown heat recovery (ECM 2). Multiple standard or custom tailored reports/charts could be generated to serve any purpose.

The performance of the equipment and systems is reported to the plant on monthly basis and used to generate invoices for each utility. An overall report is prepared on annual basis and provided to the plant for performance evaluation.

CONCLUSION

By implementing the above nine energy conservation measures, the plant captured $566,000 (at 2002 utility costs) in annual energy savings, and reduced the total carbon emission to the environment with 1370 tons annually.

The identified opportunities to save fuel, steam and condensate are summarized in Table 13. When implemented they lead to 27% overall steam system efficiency improvement.

The installation of EOS made the process of data acquisition, monitoring, recording, trending, reporting and control possible not only for the plant operation and maintenance personnel, but also accessible from a standard web browser for a control and management prospective.

BIBLIOGRAPHY

1. Design and operating data from 2002 to 2005 client's facilities.
2. Turner, W.C. *Energy Management Handbook*, 1992.
3. Armstrong International Inc., *Steam Conservation Guidelines for Condensate Drainage*, Handbook N-101, 2005.

Solid–Sus

Steam Turbines

Michael Burke Wood
Ministry of Energy, Al Dasmah, Kuwait

Abstract
This article reviews the history of steam turbine development for power generation and marine propulsion. It describes factors affecting thermodynamic and fluid dynamic efficiency and mechanical design principles. Alternative design technologies are described and some comparisons are given. Steam turbines for various applications and probable future developments are described.

INTRODUCTION

The steam turbine has been the predominant prime mover for electricity generation for a century. Its efficiency depends on steam conditions, steam cycle, and turbine fluid dynamic efficiency. Materials properties and mechanical design determine possible steam conditions and turbine rating. Fluid dynamic efficiency is determined by blade and steam path design. These factors and their interaction are described. The changing role of the steam turbine to combined cycle applications, with uncertainty regarding the scope for fossil-fired applications using advanced steam conditions and for new nuclear power plants, is discussed.

HISTORY

Hero of Alexandria's "aoelipile" demonstrated steam-driven rotary motion in 200 B.C. It was driven by the thrust of two tangential outlet pipes and was therefore a reaction turbine. In Italy in 1629, Branca proposed fixed steam jets impinging on rotating blades—the impulse turbine. Both machines were without practical application. Until 1900, motive power was dominated by steam engines. In 1784, James Watt dismissed the turbine as a competitor. He reasoned that steam at boiler pressure would escape to the atmosphere at 2000 ft/s; and because blade speeds of half that were required to extract a useful proportion of its energy, it was impossible to construct a machine to withstand the resultant centrifugal loading. In 1837 in the United States, William Avery used Hero's concept to build machines up to 5 ft in diameter, with peripheral speeds of 900 ft/s, to drive woodworking machinery. They were abandoned because of control difficulties and unreliability. Gustav de Laval in Sweden used a similar device to power a centrifuge, then in the 1890s turned to single-stage impulse turbines. These used a convergent-divergent nozzle to fully expand the steam and had rotor speeds of up to 30,000 rpm. The ascendancy of the steam turbine began when Charles Parsons solved the problem of high velocities by applying pressure compounding to divide pressure drop between a large number of reaction stages along the axis of the turbine. In a year from 1884, he patented and demonstrated the first practical turbine generator. In 1886 in France, Professor Rateau pressure compounded the impulse turbine. In the United States, electrification proceeded rapidly, and in the mid-1990s, George Westinghouse acquired U.S. rights to Parsons' patent and installed the first commercial turbines in 1899. Charles Gordon Curtis addressed the steam velocity problem by velocity compounding; i.e., directing steam from a single row of nozzles onto several rows of moving impulse blades with intervening fixed guide blades. Vertical-axis 5-MW Curtis Turbine Generators were installed in Chicago in 1903. In the United Kingdom, the most rapid development was marine. In 1897, Parsons raced through a fleet review by Queen Victoria in his launch, Turbinia. It was capable of 34 knots, and outpaced the fastest Royal Navy patrol boats. The implications were clear, and by 1905, turbine power was exclusively selected for British warships. It was also used for liners from 1906 until the end of their era.

After 1910, development was rapid, aided by the acceptance of alternating current, which permitted efficient blade design in turbines direct coupled to generators operating at 3600 rpm (60 Hz) in the United States and 3000 rpm (50 Hz) elsewhere. By the 1920s, efficiency was improved by condensers, superheating, and feed heating. In marine applications, multiple reduction gearing allowed diameter and speed to be optimized, resulting in major increases in efficiency. Impulse turbines had the advantages of high power density, smaller rotors, and greater tolerance of rapid temperature changes. They were affected by blade vibration but came to dominate this field after World War II. Reheat was adopted generally at this time, and turbine generator size and efficiency continued to increase. Supercritical steam cycles with steam pressure and temperature conditions of 340 bars/650°C and double

Keywords: Steam turbine; Rankine cycle; Blades; Losses; Mechanical design.

reheat were introduced in the 1960s. Materials difficulties caused reduction to 240 bars/560°C, typical of supercritical units now in service in the United States. Steam turbines operating at up to 250 bars/600°C are used in Europe and Japan. Nuclear reactors require low-speed, wet-steam turbines of up to 1500 MW operating at 75 bars/295°C. In the last decade, combined cycles have become a major steam turbine application. In the marine field, the liners were eliminated by airlines soon after World War II. Rising fuel costs and large, very efficient diesel engines eliminated steam turbines from merchant ships by 1980, despite attempts to compete using very advanced 210 bars/600°C, 14,000-rpm designs. Naval propulsion is now predominantly gas turbine and diesel except for nuclear powered capital ships and submarines.

STEAM CYCLES

The Carnot cycle, shown in Fig. 1, is the most efficient fluid power cycle. It illustrates principles and the reasons for more complex cycles. Processes are as follows: 1–2, reversible isothermal heating in the boiler; 2–3, isentropic expansion in the turbine, 3–4, reversible isothermal condensation; and 4–1, isentropic compression. The area under 1–2 represents heat from the boiler, and that under 3–4 represents heat rejected to the condenser. The difference between the two areas is the work produced. Carnot Efficiency is defined as $(T_{max} - T_{min})/T_{max}$. In a steam cycle, the condenser cooling medium fixes T_{min}; and increasing T_{max}, the temperature at which heat is supplied, maximizes efficiency. This cycle is impractical because isothermal heat supply is only possible in the two-phase region, limiting maximum cycle temperature to below the critical temperature of 374°C. Expansion of saturated steam gives poor steam quality at turbine exhaust, which would cause erosion of the turbine blades. A two-phase compressor is impractical. These difficulties are overcome by the Rankine cycle, shown in Fig. 2, in which heat is supplied at constant pressure from 2 to 3 in successive economizing, evaporating, and superheating. Superheating increases the average heat supply temperature and steam quality at turbine exhaust. Complete condensation from 4 to 1 necessitates only a liquid phase boiler feed pump. Fig. 2 also indicates the effect of irreversibility due to the fluid friction and heat loss in a real cycle. Fuel costs are many times capital costs; and therefore, a complex cycle is justified to maximize efficiency. The values of maximum cycle temperature permitted by creep properties of boiler and turbine materials result in considerable wetness after expansion. For a given maximum temperature, raising the pressure—hence, also raising boiling temperature and average heat supply temperature—increases efficiency. It has the disadvantage of increasing leakage losses and displacing the expansion line to the left, further increasing exhaust wetness. To reduce wetness, steam is returned to the boiler after partial expansion and reheated, as shown in Fig. 3. This further increases average heat supply temperature and efficiency by 4%–5%. Two stages of reheat are sometimes used in supercritical plants. In the economizer, heat is added at low temperature. Efficiency is therefore further improved in the Regenerative Rankine Cycle (Figs. 4 and 5) by feed heaters, heated by steam

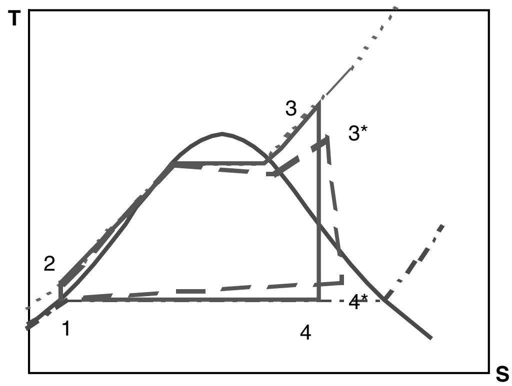


Fig. 2 Rankine cycle.

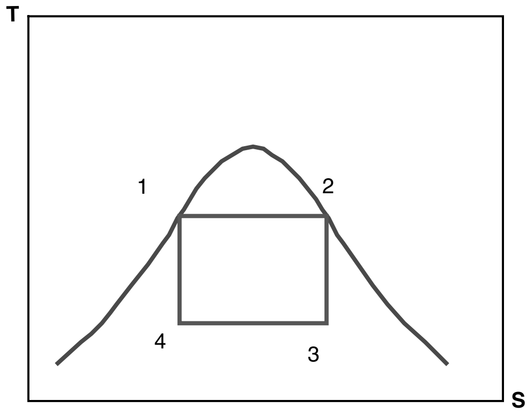


Fig. 1 Carnot vapour cycle.

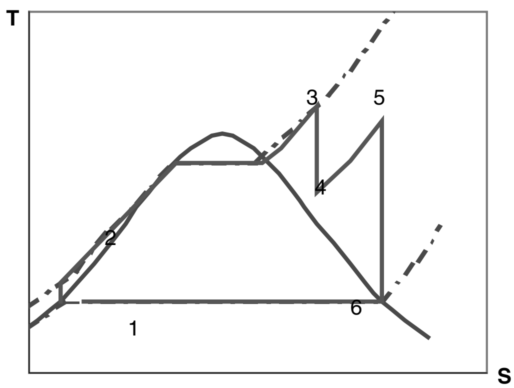


Fig. 3 Reheat rankine cycle.

Solid–Sus

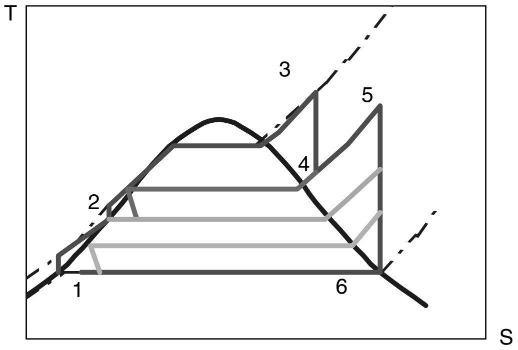


Fig. 4 Regenerative reheat rankine cycle.

extracted from the turbine. Fig. 5 is simplified. In real plants, up to 10 feed heaters are used. The overall energy efficiency of the steam cycle is given by the heat rate, defined as the heat supplied by the boiler (in Btu) divided by the generator output (in MWhr).

The use of a two-phase working fluid has major advantages. The efficiency of energy conversion in heat engines is a function of the pressure ratio. In petrol engines, this is about 12:1; in a diesel engine, 18:1; and in a typical gas turbine, 20:1. In the steam cycle, because of condensing, exhaust pressure is very low, and this ratio is at least 4000:1. The power required for the pump or compressor in the cycle is related to the change in volume. In the Rankine cycle, pumping takes place in the liquid phase; therefore, power requirements are 2 orders less than in a cycle using air as a working fluid. Energy release per unit mass of working fluid is very high, increasing the power density of the heat engine.

STEAM PATH

Steam is expanded in fixed blades or nozzles attached to the casing, and its enthalpy is converted to kinetic energy then to mechanical energy by moving blades attached to the rotor. In the United States, blades are referred to as buckets. The efficiency of the conversion is the ratio of the actual turbine work in process 3* to 4* (Fig. 2) to the isentropic turbine work in process 3–4. This ratio depends on turbine design and losses but is greater than 90% in modern machines. A row of fixed blades and the following row of moving blades is called a stage. The combination of rotors and casings between bearings is a cylinder. This may include two flow paths. Specific volume, and therefore required steam path flow area, increases by a factor of 2000 from inlet to exhaust for subcritical turbines, and by a factor of 3000 for supercritical turbines. This is accommodated by increasing the blade height and rotor diameter from stage to stage through the steam path from the inlet through high- (HP), intermediate- (IP), and low-pressure (LP) cylinders; and, if necessary, by double-flow IP and several double-flow LP cylinders.

Blades and Seals

In a pure impulse stage, pressure drop occurs across the fixed blades and enthalpy is converted to kinetic energy.

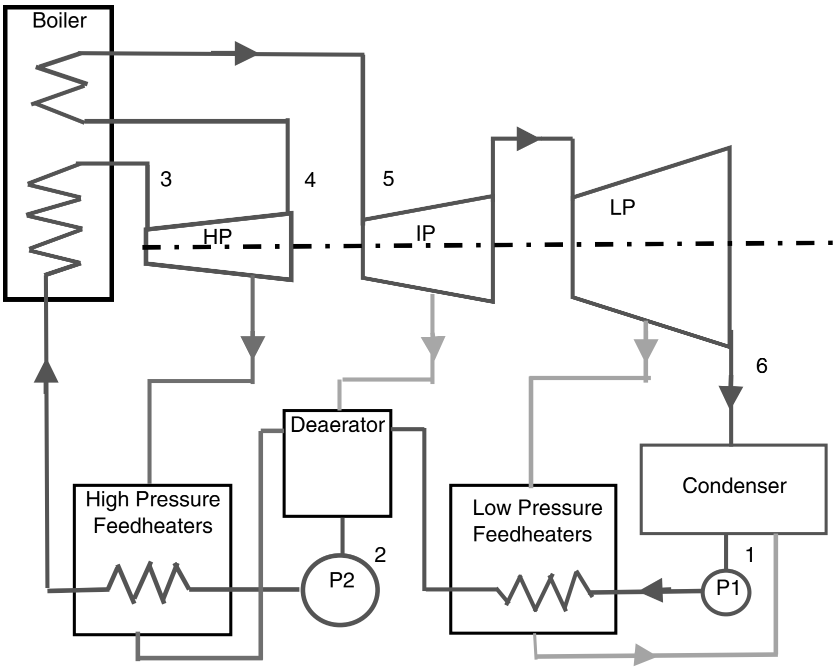


Fig. 5 The regenerative reheat rankine cycle.

Moving blades change the direction of the steam so that energy is transferred to the rotor entirely by change of momentum. Only Hero's concept is pure reaction. All blades have an impulse component. In reaction stages, the flow path between rotating blades is shaped to cause further pressure drop, expansion, and kinetic energy release, which is transferred to the rotor by reaction. The percentage of enthalpy drop across the moving blades is the degree of reaction. Also, in any blade, centrifugal effects in the space between fixed and moving blades increase pressure drop across the latter, and therefore the degree of reaction with radius. All blades have some degree of reaction, varying from a minimum of 5% in "impulse" to 50% in "reaction" designs. These differences have a marked effect on HP and IP design. In impulse design, stationary blades are arranged in rigid diaphragms with a labyrinth seal to the rotor. The diameter of the seal, and therefore leakage area, is minimized. Between seals, the rotor diameter is increased to form a wheel to which the rotating blades are attached. This construction is called "diaphragm and wheel." The pressure drop across moving blades is small, but sufficient to require a tip seal. In reaction design, it is greater and requires higher tip seal duty per moving blade row. It also causes an axial thrust on the rotor; therefore, annular area between the fixed and rotating blade seals must be a minimum, precluding diaphragm and wheel construction. Rotors of cylindrical or slightly tapering conical form, referred to as drum rotors, are used. Thrust is balanced either by a large annular rotor surface subject to steam pressure, a dummy piston; or by combining opposed flows in the same cylinder. For the same enthalpy drop, a reaction turbine requires about twice as many stages as an impulse turbine, but the two concepts have remained competitive through their development histories to present, both having very high efficiencies. Reaction stage efficiencies are higher, but differences in overall turbine efficiency are evidently small, and could only be identified from proprietary information. Diaphragm and wheel construction has the advantages of compactness and lower transient thermal stresses. In LP blades, differences between the two concepts diminish. The last stage can produce up to 10% of total turbine output and is most important in turbine efficiency. Steam volume is large and blades are up to 1.2 M long in 3000 rpm machines. There is a large change in blade speed from root to tip, whilst the axial steam flow is more nearly constant. To align with the resultant steam flow, the angle of the blade is twisted from near axial at the root to near tangential at the tip. Increases in the degree of reaction from root to tip due to centrifugal effects are large, and a marked change in blade profile is required. Blades with these features are known as vortex blades. Full analysis of LP blades is complex and has only become possible in the last two decades with the advent of computational fluid dynamics, correlated with experiments. Their optimization is one of the major recent advances in turbine design.

Losses

Turbine efficiency is determined by fluid dynamic, leakage, wetness, and leaving losses. Fluid dynamic losses due to surface friction and secondary flows are small in modern HP but more significant in impulse stages and LP blades, due to higher steam velocities. Curved nozzles, which lean in the tangential direction, have been developed to reduce secondary flows. Leakage losses due to steam bypassing the tips of moving and stationary blades are more significant in HP, particularly in reaction stages. In LP stages, when steam is wet, water droplet velocities are less than steam velocities and cause drag on, and erosion of, moving blades. Stage efficiency reduces by approximately 1% for 1% additional wetness. Within the steam path, kinetic energy leaving one stage is used in the next, but that leaving the last stage is lost in the condenser. "Leaving loss" is proportional to the square of exhaust velocity. The greater the diameter of the last stage blades and the number of exhaust flows (and therefore the greater the exhaust area), the lower the leaving loss. This is one of the most important issues in determining turbine size, cost, and efficiency. The design of the duct between the turbine exhaust and the condenser—the exhaust hood—is also important. Modern designs can produce a backpressure below condenser pressure, so increasing turbine output.

MECHANICAL DESIGN

Turbines must be designed for temperatures of up to 600°C; pressures of up to 300 bars; transient thermal stresses (mainly during startup and shutdown); and fluid dynamically, and mechanically induced vibration, corrosion, and erosion. In HP cylinders, centrifugal loading essentially depends on peripheral velocity, which is defined by blade efficiency. Optimum designs for pressure containment and minimum thermal stresses are obtained by minimizing rotor diameter and casing thickness, and therefore by achieving high rotational speed. For power generation, gearing is impractical; therefore, optimum rotational speed is either 3000 or 3600 rpm, depending on grid frequency. Impulse designs have advantages with regard to thermal stresses. They have a smaller rotor body, which reduces maximum values, which also do not occur at stress concentrations due to blade attachment details. For LP rotors with large blades, centrifugal loading depends on speed of rotation; therefore, 1500 or 1800 rpm allows the largest blades and exhaust area. For these reasons, until about 1970, large turbines were optimized by cross compounding with a high-speed HP/IP shaft and a low-speed LP shaft, requiring two generators and very large turbine halls. Developments in LP blade design now enable much more compact machines of over 800 MW to be built with 2 LP cylinders,

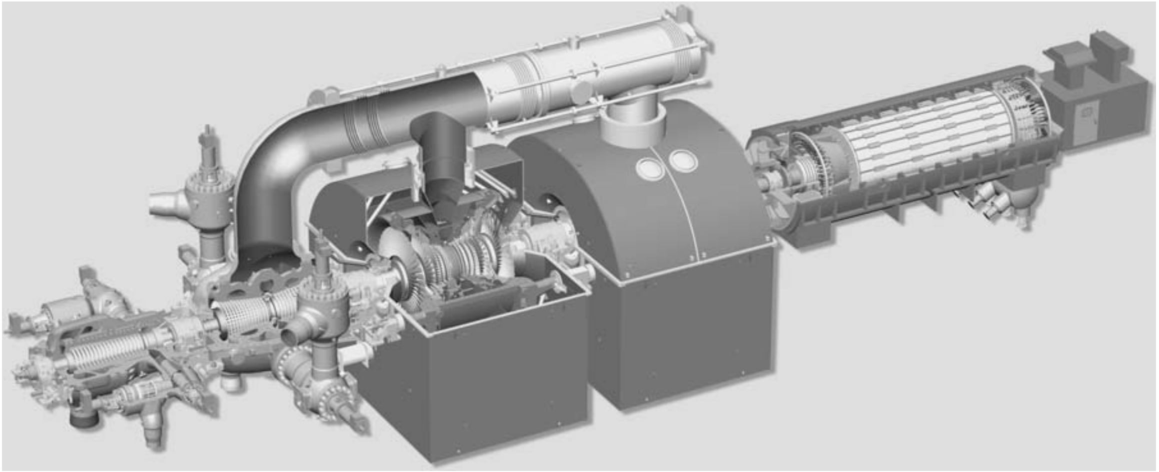


Fig. 6 Arrangement of 4 cylinder up to 1200 MW, 300 bar, 600°C, reheat steam turbine. (Courtesty of Siemens AG.)

and up to 1100 MW with 3 LP cylinders, both using a single high-speed shaft.

Creep properties of high-temperature sections, tensile strength of LP blades and rotors, and toughness to prevent failure of rotating components are the most important material issues. Turbines are large capital investments and must operate reliably for more than 30 years—200,000 h with minimum possible outages. In the past, design was evolutionary; but modern practice is increasingly influenced by engineering analysis of thermal, mechanical, and chemical operating and fault conditions; and by materials development. Extremely high reliability has been achieved. Overhaul intervals approaching 10 years are possible, justified by condition monitoring, and machines' lives extended to more than 40 years. Figs. 6 and 7 show a large modern reaction turbine with a single-flow HP cylinder, a double-flow IP, and two double-flow LP cylinders. Condensers are beneath the LP cylinders. This configuration is used for ratings to over 800 MW and supercritical steam conditions approaching 600°C and 300 bars with a single reheat. Rigid couplings connect all rotors, including the generator. One LP casing is anchored to the foundations to minimize expansion movement relative to the condensers. Other casings are supported on sliding feet to allow free longitudinal expansion. A single thrust bearing is positioned to minimize axial differential expansion between the rotors and casings. Bearings are set at different levels, matching the deflected forms of the rotors to minimize rotating bending stresses. High pressure and IP cylinders have inner and outer casings. The inner casing, which carries the stationary blades, contains the difference between the stage pressure and the cylinder outlet pressure. The outer casing contains the outlet pressure. This allows the casings to be thinner, minimizing

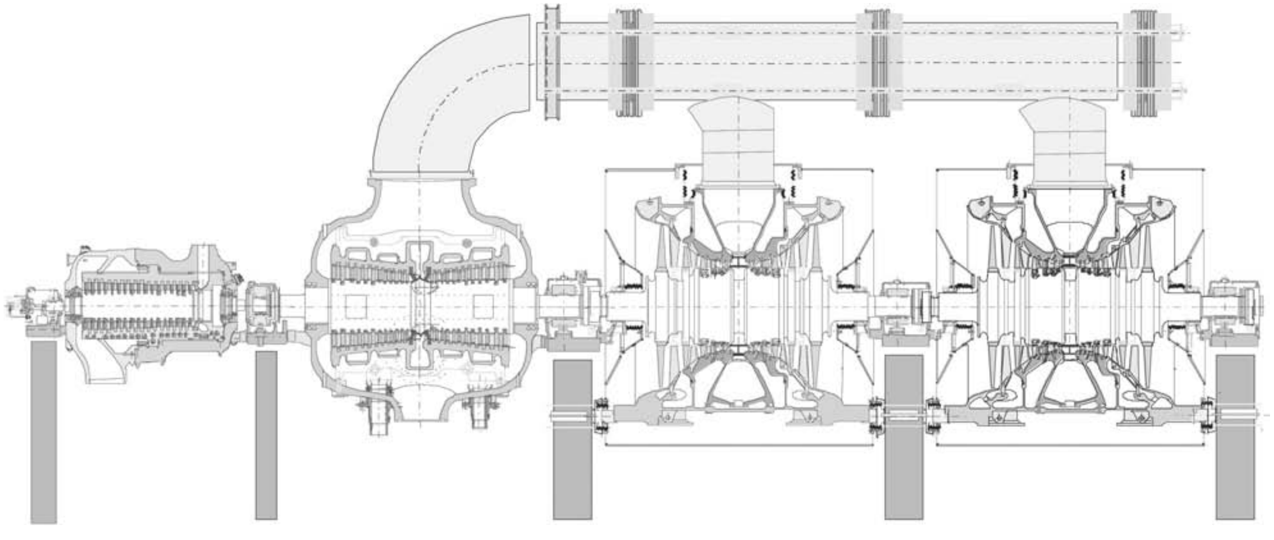


Fig. 7 Section of 4 cylinder up to 1200 MW, 300 bar 600°C, reheat steam turbine. (Courtesty of Siemens AG.)

thermal stresses. In other designs, pressure loading on the casing is further reduced by use of an impulse first stage and a separate inlet nozzle box containing the large pressure drop across the first-stage nozzle. This also reduces the number of stages at some cost in efficiency. Most turbines have horizontal bolted joints in both inner and outer casings to permit assembly and maintenance. Massive flanges with a very high ratio of bolt to flange area are required. Because the casing heats up more rapidly than the flanges during start up, rates must be limited to minimize thermal stress, distortion, and thermal fatigue. In impulse designs, diaphragms reduce leakage caused by casing distortion. In some designs, inner casing bolts are replaced by shrunk-on rings requiring complex assembly and disassembly. High pressure and IP Casings are steel castings; generally, 1%–2% chromium alloys for sub-critical turbines and 9%–12% Cr for supercritical turbines. Low-pressure casings have an inner nodular cast-iron casting carrying the blades, supported in a fabricated steel outer casing including the exhaust hoods. Fig. 8 shows assembly of an LP cylinder. High pressure and IP rotors are one-piece forgings in similar materials to those of the casings, alloyed to maximize toughness. Low-pressure rotors for fossil-fired applications are also normally one-piece forgings. Those for wet steam nuclear turbines comprise a shaft with shrunk-on disks to carry the blades. Compressive stresses in this design increase its resistance to stress corrosion cracking. Because of the high centrifugal loading due to very long blades, low-nickel, chrome, vanadium alloys with very high strength and toughness are used.

Blades are generally 12% chromium or higher steel alloy forgings. They are attached to the rotor by a mechanical connection consisting of a lobed male detail on the blade and an identical female detail on the rotor. The design most commonly used for LP blades is shown in Fig. 9. Similar concepts oriented circumferentially are used for attachment to drum rotors. Multiple pinning through a forked blade root and the disk, over which it is fitted, is also used. Mechanical design of last-stage blades is one of the major factors controlling turbine output and configuration. Length is determined by centrifugal loading. Titanium is now used to increase the material-strength-to-weight ratio. Blade vibration and avoidance of high-cycle fatigue are also major considerations. In the past, blades were rigidly connected in groups to control vibration. Two design approaches are now in use. The blade on the left of Fig. 9 is free standing, designed to ensure that no modes of vibration are excited in operation. The alternative is continuous circumferential connection, suppressing some modes and providing damping. Fatigue of inlet HP blades is also important when steam is admitted over only part of the circumference, and steam forces on the blade vary as it enters and leaves the admission zone.

Sealing is by "labyrinth packings." These consist of several thin metal disks, spaced axially and mounted on the casing with minimum clearances from the shaft. Resistance to leakage is provided by the small clearance gaps in series, and by creating a tortuous flow path. Materials are chosen to prevent damage to the shaft,

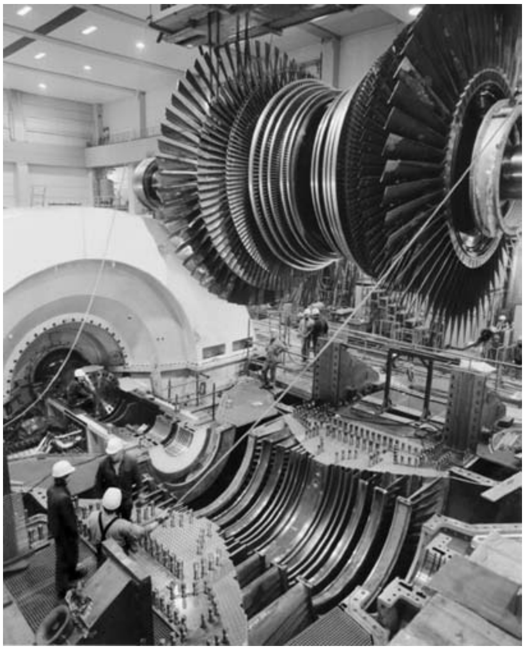


Fig. 8 Low pressure cylinder assembly. (Courtesty of Siemens AG.)

Fig. 9 Low pressure rotor. (Courtesty of Siemens AG.)

should a rub occur. Glands, seals between casings and rotors, are required to prevent outward steam leakage from HP and IP cylinders and inward air leakage to LP cylinders, which would reduce condenser vacuum. These consist of several labyrinth packings in series. Steam at IP is introduced between two packings. This controls the pressure gradient along the seal, minimizing leakage. Leakage increases the heat rate, and is minimized by intermediate leak-off paths routed to the feed system. Steam which passes through the whole assembly is captured by a gland steam condenser, maintained at subatmospheric pressure.

Hydrodynamic white metal journal bearings are used. For LP cylinders, bearings with length greater than diameter are required to obtain suitable bearing pressures. When the turbine is shut down, HP and IP rotors remain at temperatures at which permanent creep bending under self-weight would occur. To prevent this, it is essential to rotate the turbine slowly until it cools sufficiently. This is referred to as "barring." Under these conditions, the hydrodynamic oil film is not developed, and alternative pressure lubrication is normally provided.

TURBINE CONTROL

Valves in the steam chests, which can be seen in Figs. 6 and 11, control speed and load. If these are operated in parallel, the turbine is "throttle governed," and steam flow and pressures through the turbine are proportional to load. The disadvantage is that there is a loss of efficiency at part load due to throttling in the valve. Operating the valves sequentially can reduce this. At part-load, when only one valve is used, throttling is less. This is "nozzle governing." Throttling losses can also reduced by operating the boiler at variable pressure; i.e., sliding pressure control, used in combined cycle steam turbines. "Bypass governing," whereby steam is admitted partway down the steam path, can be used if the boiler must operate at constant pressure. Overspeed in the event of load loss and consequent damage is prevented by rapid closure of the valves. In this situation, energy stored in steam in the reheater and feed systems must be isolated from the turbine by further rapidly acting valves.

TURBINE TYPES

The machine shown in Figs. 6 and 7 is typical of those used in a modern fossil-fired plant. The predominant fuel is coal, but lifetime fuel costs are still several times capital cost; therefore, a supercritical plant is generally built. The most common size is about 600 MW, but up to 1200 MW is possible with two LP cylinders. For smaller sizes, the configuration shown in Fig. 10 is attractive and widely used. It is suitable for loadings up to 500 MW, with subcritical steam conditions of 177 bars and up to 600°C. The cylinder on the left includes opposed-flow HP and IP turbines. Steam enters near the center, expands to the left in the HP, is reheated to approximately the same temperature, re-enters near the center, expands to the right in the IP, then exhausts to the LP via the outer casing. This configuration confines high temperatures to the center, reducing thermal stresses and allowing more rapid starting. It is more compact and economic than separate HP and IP cylinders.

Pressurised Water (PWR), Boiling Water (BWR), and Canadian deuterium uranium (CANDU) nuclear plants generate saturated steam at approximately 75 bars and 295°C. It is dried, admitted to the HP turbine, and expanded until it is about 10% wet. It is then exhausted to water separators and superheated in a steam-to-steam heat exchanger before admission to the LP turbine. Low-pressure superheating increases average heat supply temperature and efficiency and limits wetness to about 10% at exhaust to the condenser. Such steam turbines drive four-pole generators and rotate at half the speed of conventional machines. For a given output, flow areas are doubled and velocities are halved. This results in very large, multicylinder turbines but has the advantage of reducing blade erosion by water droplets in wet steam and reducing leaving loss. Machines of up to 1500 MW with three opposed-flow LPs have been built.

Because of lower heat rates, capital costs, and construction times than conventional steam plant, combined cycle gas turbine (CCGT) is now the largest application of steam turbines. Gas turbines exhaust into individual heat recovery steam generators (HRSG's). One or more supply steam to the steam turbine. If only one is used, the gas and steam turbines are arranged on a common shaft, connected by a clutch to allow the gas turbine to be started first. To obtain maximum heat recovery, up to three steam pressures are used in the HRSG, and steam is admitted to the turbine at three positions along the steam path, increasing steam flow from inlet to exhaust. No feed heaters are used, and steam extraction is only for the deaerator. The importance of last-stage blade efficiency is increased. Fig. 11 shows a typical two-cylinder machine for combined cycle application. It is capable of up to 250 MW, with steam conditions of up to 600°C and 170 bars. It is designed for installation on a flat foundation with an end-mounted condenser.

Steam turbine generators are frequently combined with process plants. This is called cogeneration. If the process requirements and generation requirements are fixed, steam is produced at a higher pressure and expanded in a back pressure turbine driving a generator, then exhausted at the pressure required for the process plant. If the steam required for the process is less than that for generation, a pass-out or extraction turbine is used, process steam is extracted from the turbine after partial expansion, and the remainder fully expanded before exhaust to the condenser.

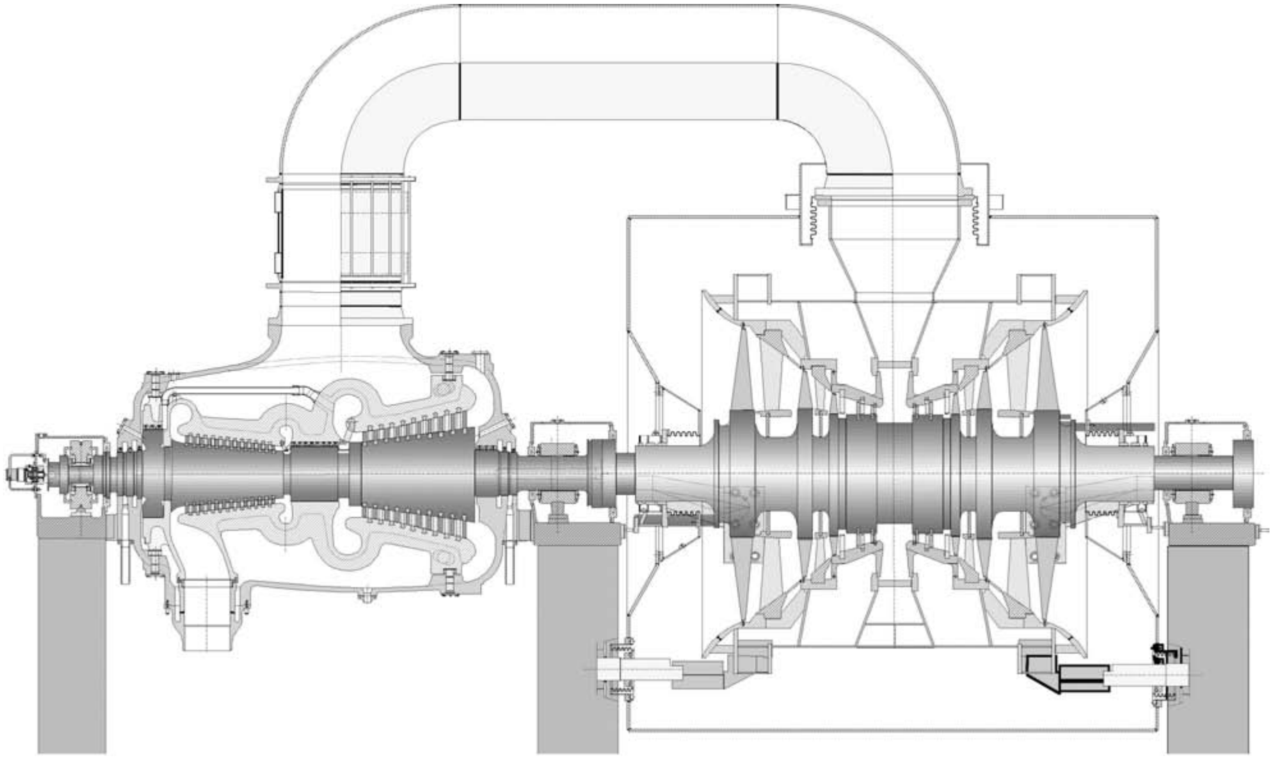


Fig. 10 Arrangement of 2 cylinder up to 500 MW, 177 bar, 600°C, reheat steam turbine. (Courtesy of Siemens AG.)

DEVELOPMENT

Steam turbine technology is thoroughly evolved, and the limits set by material properties and fluid dynamic efficiency are well defined. Applications depend on other technologies. Combined cycles require relatively simple machines with emphasis on maximum exhaust efficiency. Developments will probably be the more complex steam

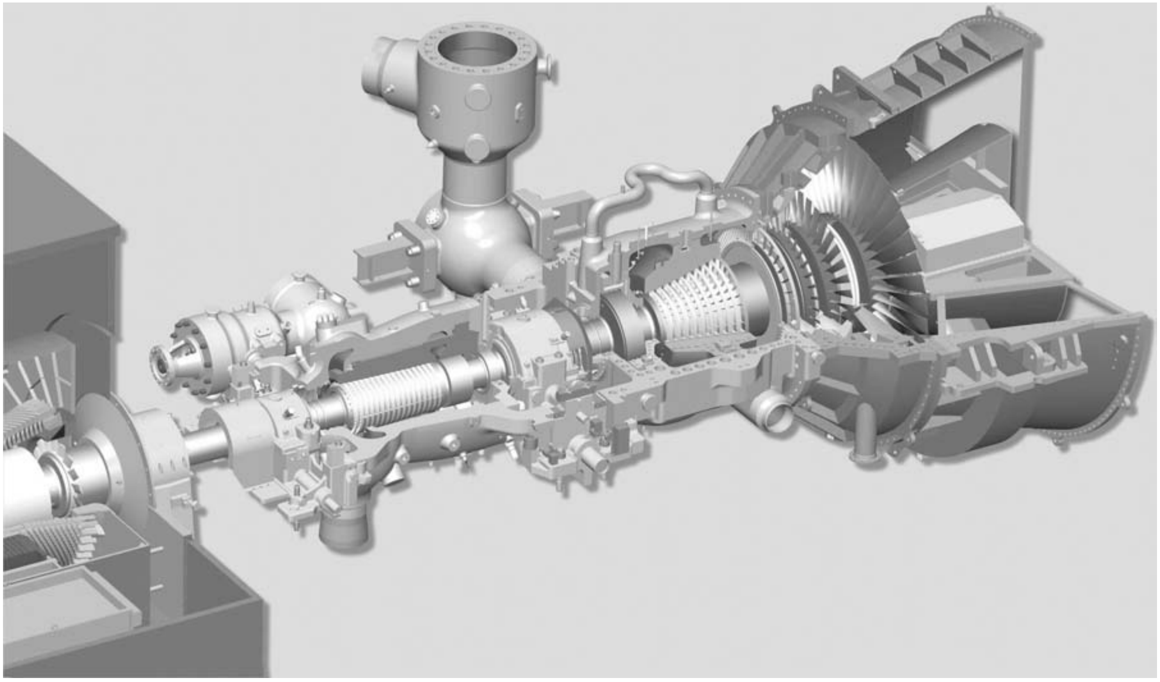


Fig. 11 Arrangement of 2 cylinder up to 250 MW, 177 bar, 600°C, reheat steam turbine, suitable for combined cycle applications. (Courtesy of Siemens AG.)

path requirements of advanced gas turbines using steam-cooled gas turbine blades and integrated gasifiers. Coal remains a major energy source, and the supercritical steam plant with exhaust conditioning is an economically and environmentally competitive technology for burning it. Related developments are supercritical steam turbines for steam conditions above 700°C, longer last-stage blades, and units with larger outputs and fewer cylinders. Steam turbines are essential for power production from water-cooled nuclear reactors, and steam conditions may increase and unit ratings may reduce in the future.

CONCLUSION

The steam turbine is an elegant concept, capable of converting the energy of steam into mechanical work solely by rotary motion. Condensation of steam enables exceptionally high pressure ratios and energy extraction. Its characteristics as a prime mover are ideal for electricity generation. A steam power plant is capable of over 40% overall energy efficiency, firing a wide range of fuels. The turbine's mechanical elegance results in low maintenance requirements and long service life. For these reasons, it has remained predominant in this application for a century. In this period, efficiency and power density have increased by factors of more than 4, and remarkably, two alternative steam turbine technologies have developed competitively throughout. It had a major role and underwent major development in marine propulsion. This ended because of changes in shipping economics and large, highly efficient diesel engines. Sole predominance in power generation is now ended by large gas turbines, but only combination with the exceptional energy recovery of the steam tubine enables efficiency over 55% to be achieved in combined cycles. These can be adapted to all fuels by the use of gasifiers. Applications of large turbines for nuclear reactor power cycles and coal-fired steam plants with potential for power outputs of up to 1700 MW will continue; but their relative importance depends on a resurgence of nuclear power in response to global warming, the availability of gas to fire CCGT, and advanced clean coal technologies, including gasifiers.

ACKNOWLEDGMENT

Siemens, A.G., for comments and provision of the illustrations included in this article.

BIBLIOGRAPHY

1. Cengel, Y.A.; Bowles, M.A. Vapour and combined power cycles, 4th Ed. In *Thermodynamics, an Engineering Approach*; McGraw-Hill: New York, 2002; 541–547, [Chapter 9].
2. Kearton, W.J. Flow of steam through impulse and impulse—reaction turbine blades. In *Steam Turbine Theory and Practice*; Pitman: London, 1966; 172–227, [Chapters 7–8].
3. El Wakil, M.M. Turbines, 1st Ed. In *Powerplant Technology*; McGraw-Hill: New York, 1984; 173–210.
4. Kehlhofer, R.H.; Warner, J.; Nielsen, H.; Bachmann, R. *Components, Combined-Cycle Gas Steam Turbine Power Plants*, 2nd Ed.; Penwell: Oklahoma, 1999; 155–188.
5. Nicholas, D.G.; Burn, R.E. Marine steam turbines at the centenary. Proceedings of First Parsons International Turbine Conference, University of Dublin, 1984 The Parsons Press: Dublin, Ireland, 1984, 17–36.
6. Mitchell, J.M.; Grant, J. Design and fluid dynamic development of steam turbine blading. Proceedings of First Parsons International Turbine Conference, University of Dublin, 1984, The Parsons Press: Dublin, Ireland, 1984, 81–91.

Sustainability Policies: Sunbelt Cities

Stephen A. Roosa
Energy Systems Group, Inc., Louisville, Kentucky, U.S.A.

Abstract
It has been suggested that "sustainability can provide a qualitative measure of the integrality and wholeness of any given system.[1] This article focuses on evidence indicating that a positive relationship may exist between the adoption of sustainability as a local goal and the rates of local energy policy adoption in Sunbelt cities. The evidence suggests that policies responding to energy and environmental issues may provide long-term solutions to achieving urban sustainability.

The fundamental question becomes "What evidence is there that cities with sustainability as a stated goal have higher rates of local energy policy adoption?" In this article, it is asserted that cities with sustainability as a local goal are more likely to adopt certain energy related policies. It is also determined that there is variability in the implementation of energy related policies in Sunbelt cities. The author concludes that Sunbelt cities with sustainability as an urban goal have higher rates of energy policy adoption when the three selected policies are studied as indicators.

INTRODUCTION

Cities have common energy concerns that impact their urban environments. Cities seeking to incorporate policies that lead to sustainability generally consider energy policy to be a critical component of their urban agendas. These concerns are manifested in urban policies and programs designed to achieve results that include reducing energy usage with the potential of improving the environment. The purpose of this research is multifold:

- Determine which Sunbelt cities have sustainability goals.
- Provide a descriptive comparison of select energy-related policies in Sunbelt cities.
- Identify the specific types of policies that are being adopted and pursued.

In this article, the concept of sustainability is defined and cities with sustainability agendas are identified. The 25 largest cities in the Sunbelt (noted in the attached tables) have been selected for consideration. This research provides an assessment of the selected cities based on the energy-related policies they have adopted. To provide evidence of policy adoption, three locally adoptable energy-related policies are considered. The three policies considered are (1) city operated energy efficiency programs, (2) local governmental energy program support, and (3) Energy Star™ program participation.

Keywords: Energy Policy; Sustainability; Local Energy Policies; Sunbelt.

URBAN SUSTAINABILITY

Sustainability is a broadly defined concept that has a variety of meanings. Urban sustainability refers to an idealized model of urban development that attempts to address concerns about urban growth, patterns of urban development, and issues that arise as urban development occurs. According to Beatley,[2] the four principles of urban sustainability include (1) the principle of urban management, (2) the principle of policy integration, (3) the principle of ecosystems thinking, and (4) the principle of cooperation and partnership. How cities choose to manage energy policy involves each of these principles. Energy policy requires management, needs to be integrated with other urban policies, impacts many ecosystems, and requires cooperation and partnership to be successfully pursued.

For the purposes of this article, sustainable development is defined as the ability of physical urban development and urban environmental impacts to sustain long term inhabitation by human and other indigenous species while providing (1) an opportunity for environmentally safe, ecologically appropriate physical development; (2) efficient use of natural resources; (3) a framework which allows improvement of the human condition and equal opportunity for current and future generations; and (4) manageable urban growth. Non-sustainable urban development is the antithesis of sustainable urban development; it implies growth that is environmentally unsafe, consumes resources inefficiently, degrades the human condition, is characterized by persistently unmanageable development, and fails to value social equity. The energy policies cities choose to adopt are critical to the success or failure of urban sustainability programs. Inappropriate use of energy can cause substantial damage to economies, the environment and create the need for corrective policies. Energy is a resource that begs to be used efficiently. Gross consumption of available energy resources can cause a tragedy of the commons, preventing future use of the resource. Urban growth can be impacted by dependence on fossil fuels.

COMPARING SUNBELT CITIES AND THEIR POLICIES

Why are Sunbelt cities and their sustainability policies particularly interesting? Sunbelt cities are significant centers for urban population growth and development in the United States, generally outpacing their non-Sunbelt counterparts. Many of the selected cities in this broadly defined region (e.g., Las Vegas and Phoenix) are among the fastest growing cities in the United States. A few are leaders in implementing innovative policies. Other cities used in this study, including Atlanta, Austin, Phoenix, Nashville, and Oklahoma City, are their state capitals, making them centers of statewide decision making. With new investment and construction, Sunbelt cities have the opportunity to select from a range of newly available technologies when growing their cities. Their policies will ultimately impact the design of the cities, their future energy usage, and the sustainability of their urban areas.

The cornucopia of policies available to urban regimes to achieve reductions in energy use might include transportation system policies, energy management programs, organizational memberships, and policies designed to improve the environment. However, this examination focuses on three selected indicators: (1) city-operated energy efficiency programs, (2) local government energy program support, and (3) Energy Star program participation. The selected policies require local initiative to implement and sustain. The energy-related policies selected have broad applicability and are available for adoption in some form by all urban areas considered in this study. The first category of policy indicators focuses on whether or not sustainability is an adopted local goal.

SUSTAINABLE DEVELOPMENT AS A POLICY GOAL

This indicator poses a fundamental question: Is sustainable development or urban sustainability a stated goal of the city? Ancient southwestern monuments such as Mesa Verde and Chaco Canyon provide testimony that cities in the southwest may have been abandoned as a result of environmental mismanagement and changes in local environmental conditions. Many southwestern mining towns grew to become boomtowns, only to go bust and ultimately become ghost towns after their resources were no longer considered exploitable. Perhaps even today, we are growing new throw-away cities. If sustainability is not an identified goal of the urban agenda, then it would seem unsurprising if it is not achieved.

The use of the language of sustainability in local policymaking is a relatively recent phenomenon with variable interpretations. To consider this indicator, it must be accepted that there are multiple definitions of sustainability and that the various definitions are subject to a wide range of interpretations. Ignoring the conundrum of interpretations, this policy indicator gauges only whether sustainable development or sustainability has become a stated urban goal. However, the contextual interpretation of the concept of sustainability when searching databases is discriminately limited for the purposes of this article. For example, if a city's only stated "sustainable" policy is to "maintain a sustainable tax base," then the term is judged to be misapplied and the application discredited.

Of the 25 cities in this study, a total of seven cities (28%) have established sustainability as a primary urban goal. These cities are Jacksonville, El Paso, Long Beach, Albuquerque, Atlanta, Mesa, and Tulsa. How cities implement sustainability policies varies. In 1985, Jacksonville initiated its "Quality of Life in Jacksonville" program.[3] The pledge of Atlanta's City Council President to "create an efficient, vibrant and sustainable city" includes a 2002 energy conservation initiative.[4] In Tulsa, urban sustainability is a primary administrative goal, strongly supported by the city administration. Long Beach is atypical in that its 2010 Citywide Strategic Plan identifies "becoming a sustainable city" as a primary strategic goal. There is a statement in the *Vision of the Mesa 2025 General Plan* to support the city "as a sustainable community in the 21st century."

Seven additional cities (28%) have identified programs to support sustainable building policies, have demonstration projects underway, or have established land use requirements to promote sustainable development. San Antonio has a program to support community revitalization with a goal structure that includes "sustaining a strong urban system."[5] The City of Tucson lists a developer-driven project for the new community of Civano to develop a model sustainable community. The goal of the Civano project is "to create a new mixed use community that attains the highest feasible standards of sustainability, resource conservation, and development of Arizona's most abundant energy resource—solar—so that it becomes an international model for sustainable growth,"[6] In Los Angeles and Austin, sustainable building programs or guidelines have been established for new construction.

While 14 cities (56%) have identified sustainability as a goal or have related sustainable development policies, the remaining 11 (44%) Sunbelt cities have not established sustainability as an urban goal. It is possible to speculate that for this set of Sunbelt cities, being perceived as having a local goal of "being sustainable" may be unimportant, not considered a priority, counter to the goals of the local regime, or under consideration but not yet implemented.

Table 1 Sustainability as a local policy

Sunbelt city	Sustainability as a local policy goal
Los Angeles, CA	Yes, sustainable building program
Houston, TX	No
Phoenix, AZ	Yes, found in land use plan
San Diego, CA	Yes, goal of Environmental Service Department
Dallas, TX	No
San Antonio, TX	Yes, specific program goal
Jacksonville, FL	Yes
Austin, TX	Yes, established sustainable building guidelines
Memphis, TN	No
Nashville/Davidson, TN	No, excluded in planning mission statement
El Paso, TX	Yes, included as goal in city mission statement
Charlotte, NC	No, focus is on "smart growth"
Fort Worth, TX	No, focus is on "smart growth"
Oklahoma City, OK	No
Tucson, AZ	Yes, Adopted Sustainable Energy Code
New Orleans, LA	No, stated as goal of utility
Las Vegas, NV	No
Long Beach, CA	Yes, sustainability is the primary urban goal
Albuquerque, NM	Yes, Sustainable Community Development
Fresno, CA	No, excluded as goal in planning mission
Virginia Beach, VA	No, excluded from vision statement
Atlanta, GA	Yes, administrative goal of Mayor's office
Mesa, AZ	Yes, included in General Plan for 2025
Tulsa, OK	Yes, administrative goal of Mayor's office
Miami, FL	Yes

Among these cities are Houston, Dallas, Fresno, and Las Vegas. Table 1 provides a summary of cities and indicates which have sustainability agendas.

Other cities in this category, including Charlotte and Fort Worth, have programs with a focused development policy effort based on a "smart growth" agenda. However, policies associated with achieving smart growth agendas are not necessarily sustainable development initiatives. Having identified cities with sustainability as a local goal, three types of energy-related policies will be considered.

CITY OPERATED ENERGY EFFICIENCY PROGRAMS

Cities are owners of many and varied municipal facilities. The types of buildings owned by cities include courthouses, office buildings, educational facilities, fire and police stations, sewage treatment facilities, emergency action and preparedness centers, libraries, public health facilities, training facilities, and subsidized housing. These facilities collectively consume significant amounts of energy. Urban regimes and their administrators may view energy use as a matter of serious concern worthy of attention, as an uncontrollable overhead expense, as an unavoidable but manageable cost, as a concern only if publicly scrutinized, or as inconsequential.

The idea of civic engagement and the ethic of institutional stewardship have been linked to improving sustainability.[8] It seems logical that city governments adopting sustainable policies would be concerned with the costs and impacts of energy in their buildings and facilities. This policy indicator gauges whether the city government has an internal energy efficiency program. The actions taken by local administrations would likely be manifested in policies that support energy efficiency improvements such as the installation of energy-saving technologies, building envelope and architectural improvements, equipment replacement, and adoption of building standards among others. In order to implement facility improvements, partnerships such as performance contracts might be considered useful.[7]

This energy policy indicator provides a gauge of the importance of sustainability to local policy planners. Does the city government feel energy conservation and energy efficiency in its own buildings is important enough to warrant concerted effort? This research indicates that the administrations of most Sunbelt cities feel that energy management programs are important. In this sample of 25 Sunbelt cities, 19 (76%) have initiatives to improve energy efficiency in public buildings, while only six (24%) lacked such programs. What is striking is the wide range of approaches that cities have chosen to employ. Sample efforts implemented by Sunbelt cities include the following:

- Having a departmental division in city government for Energy Conservation and Management (e.g., San Diego).
- Hiring a City Energy Manager and implementing recommended improvements to manage and reduce energy use.

- Establishing a written energy policy for government owned buildings.
- Requiring energy assessment surveys of city-owned buildings to determine economically appropriate actions and alternatives to reduce energy use.
- Mandating the use of "Green Building" construction techniques or incorporating standards such as those required by Leadership in Energy and Environmental Design (LEED) for new construction.
- Participating in packaged programs such as the U.S. Department of Energy's (USDOEs) Rebuild America Program.
- Using energy-saving performance contracts (ESPC) as a vehicle for facility improvements with third party financing, subsidized by energy savings and cost avoidance.

The most popular policy effort among the sampled cities (with seven cities participating) was found to be the adoption of the principles and requirements of the USDOE-sponsored Rebuild America program. Rebuild America is a "network of community-driven voluntary partnerships that foster energy efficiency and renewable energy in commercial, government and public housing programs" that "works to overcome market barriers that inhibit the use of the best technologies."[9] This program is geared toward policies that reduce facility energy use while lowering the costs of energy. Among those participating in Rebuild America are the four largest Sunbelt cities (i.e., Los Angeles, Houston, Phoenix, and San Diego).

Phoenix has budgeted over a million dollars annually through 2005 to directly fund capital-intensive energy conservation improvements. Dallas, Austin, and Long Beach have adopted Green Building or LEED construction standards for city-owned buildings. San Diego's Environmental Services Operations Station administration building has carports in its parking lot with rooftop photovoltaic panels that generate 91,500 kWh per year.[10]

With over two-thirds of the city below sea level, New Orleans has concerns about rising sea levels, which threaten to displace its urban residents. As a result, in October 2001, New Orleans adopted a unique policy to reduce the threat of global warming. Greenhouse gas emissions have been profiled and municipal emission reduction targets have been mandated through 2015. Their research revealed that municipal buildings were responsible for approximately 35% of the CO_2 emissions released from municipal operations. Mitigation measures, justified by energy savings, were implemented in a number of buildings including City Hall, the court complexes, the public library, police headquarters, the airport, and others. These measures include mechanical system upgrades, installation of energy efficient lighting systems, tree planting, installation of LED new traffic signals, establishment of building energy codes for city buildings, and measures to reduce the urban heat island effect. Despite this epochal and precedent setting policy initiative, it is obvious that the actions of one city will not resolve the problems associated with global warming.

Atlanta's Energy Conservation Program exemplifies those programs that offer tangible and measurable financial returns. Atlanta's program included efforts to schedule policy workshops, perform utility rate assessments for over 600 municipal accounts, perform energy audits, and develop an internal employee energy conservation program. Within one year after initiating the program, the city had projected savings of nearly $500,000 from these initiatives.[4] In addition, the city has established a policy goal to reduce energy consumption by an additional 10% by 2010 and has appointed an Energy Conservation Coordinator.[4]

LOCAL GOVERNMENTAL PROGRAM SUPPORT

This policy indicator gauges policy support by the city government for local energy conservation, energy efficiency, and alternative energy programs. This measure asks if the local governments are active in promoting energy related programs within their respective communities.

Local energy conservation efforts can be supported by citizen actions, organizational support, corporations, utilities, local governments, other governmental bodies, or by other means. Local participation and involvement are central to the idea of sustainable cities.[1] Local governments have the primary economic means and leadership infrastructure to direct the orientation of community energy policies should they desire to assume such a role.

Among the local government entities, 14 of the 25 cities in the sample offer some sort of policy or program to support energy conservation or alternative energy, or to provide incentives for complementary technologies for new construction in their local communities. An approach that is used by seven cities is to provide financial incentives or support for community projects involving new building construction that incorporates green building technologies. The City of San Antonio has established the Metropolitan Partnership for Energy, a partnership of the city government and the community at large. The partnership has established an energy council, educational programs, facility and infrastructure improvements, equipment conservation measures, fleet conservation standards, and procurement requirements. Tucson and Las Vegas are among those Sunbelt cities that have adopted building code requirements for energy-efficient construction.

Other cities are less committed and subsidize less extensive programs. While the City of El Paso has a program, perusal of the city budget indicates that only $7500 is allocated annually. Fresno's energy policy provides for weatherization assistance for the homes of senior citizens. There are 11 cities among those sampled that lack any active energy conservation program, alternative energy policy, or support for similar local initiatives.

Combining the results from researching local governmental energy policies or programs, it was determined that (1) a total of 14 cities (56%) meet the requirements of this policy measure and have established policies or programs supported by local government, and (2) eleven cities (44%) lack policies supported by the local government.

ENERGY STAR^{TM} PARTNER

This indicator asks whether or not the city participates as a member in the Energy Star^{TM} program. Energy Star^{TM} partners include manufacturers, retailers, utilities, builders, and governments. While partnership is voluntary, there are commitments to which members must agree. To become a partner, organizations must (1) sign a memorandum of partnership committing the organization to continuous improvement of energy efficiency; (2) measure, track and benchmark energy performance; (3) develop and implement a plan to improve energy performance; and (4) educate staff and the public about the partnership and achievements of the program.[11] For urban governments, the opportunity as an Energy Star^{TM} partner is to use the label to support equipment purchasing decisions, to improve energy planning strategies, and to make better decisions concerning energy-related facility improvements.

Energy Star^{TM} is a voluntary labeling program started in 1992 that is cosponsored jointly by the U.S. Environmental Protection Agency (USEPA) and the USDOE. The focus of the labeling program concerns buildings and the energy-consuming equipment within them. Office products, mechanical equipment, lighting systems, electronics, appliances, and other products are labeled, indicating that they can be promoted as being energy efficient. The Energy Star^{TM} label has been extended to include new construction including homes, commercial structures and industrial buildings. According to its website, "through its partnerships with more than 7000 private and public sector organizations, Energy Star^{TM} delivers the technical information and tools that organizations and consumers need to choose energy-efficient solutions and best management practices."[11] Energy Star^{TM} also offers a building energy performance rating system which has been used for over 10,000 buildings throughout the United States. By leveraging private and governmental partnerships, Energy Star^{TM} has proven to be among the most cost-effective programs sponsored by the U.S. government.

For policymakers in other cities, meeting the partnership requirements might be viewed as being too costly to support and implement. A commitment to specify energy-efficient equipment might be associated with higher initial costs. A fulltime energy engineer might be required to baseline energy usage targets and establish goals. Partnership requirements might also be viewed as potentially intrusive for city administrations who consider it politically ill-advised or otherwise undesirable to advertise their ever-increasing energy costs. City administrations agreeing to measure and track energy performance, implement a plan, and improve energy performance may be subject to public scrutiny should they fail to meet objectives. As a result, some local administrators may not adopt the Energy Star^{TM} program due to the perceived potential for political risk.

Due perhaps to these and other considerations, only 10 of the 25 sampled Sunbelt cities (40%) have become Energy Star^{TM} partners. These cities are Los Angeles, Houston, San Diego, Dallas, Fort Worth, Tucson, Las Vegas, Albuquerque, Atlanta, and Miami. Consider the case of Mesa, which is among those cities that has established sustainability as a primary urban goal, yet has not chosen to be an Energy Star^{TM} partner. On the other hand, Las Vegas, Houston, Fort Worth, and Dallas are examples of cities that have not adopted sustainability goals, but happen to be Energy Star^{TM} partners.

Table 2 provides a summary of the selected policy indicators and identifies which Sunbelt cities have adopted each of the policies considered in this assessment.

FINDINGS

To determine if cities with sustainability as a local goal tend to adopt more energy-related policies, the Sunbelt cities are considered in two groups. It was stated that 14 (56%) of the Sunbelt cities have identified sustainability as a goal or have related sustainable development policies. These cities have adopted an average of 2.07 of the three energy-related policies that were considered. The remaining 11 (44%) Sunbelt cities have not established sustainability as an urban goal. These cities have adopted an average of only 1.18 of the three selected policies.

There are five cities that have adopted sustainability as a local goal (Los Angeles, San Diego, Tucson, Albuquerque, and Atlanta) and have also adopted all three of the considered energy-related policies. In addition, there are two cities that have not adopted sustainability as a local goal (Houston and Las Vegas) that have adopted all three of the considered policies. Far more cities from the sample that have adopted sustainability as a local goal have adopted all three energy related policies. Alternatively, three cities that have not adopted sustainability as a local goal have not chosen to adopt any of the three considered policies (Charolette, Virginia Beach, and Oklahoma City).

This evidence suggests that cities that have identified sustainability as a goal or have implemented related sustainable development policies are more likely to adopt energy-related policies than those that do not.

CONCLUSION

In this research, it was found that the majority of Sunbelt cities have adopted sustainability as an urban goal. It was

Table 2 Energy policy indicators

Sunbelt city	Policies for city buildings	Sponsored by local government	U.S. DoE energy star partner	Total policies
Los Angeles, CA	Yes	Green Building Initiative, Green LA	Yes	3
Houston, TX	Yes	Rebuild America, LEED Program	Yes	3
Phoenix, AZ	Yes	Yes, capital improvement projects	No	2
San Diego, CA	Yes	Green Building, Rebuild America	Yes	3
Dallas, TX	Yes	None	Yes	2
San Antonio, TX	Yes	Metro Partnership for Energy	No	2
Jacksonville, FL	Yes	None	No	1
Austin, TX	Yes	Green Building Program	No	2
Memphis, TN	Yes	None	No	1
Nashville/Davidson, TN	Yes	None	No	1
El Paso, TX	No	Yes	No	1
Charlotte, NC	No	None	No	0
Fort Worth, TX	Yes	None	Yes	2
Oklahoma City, OK	No	None	No	0
Tucson, AZ	Yes	Model Energy Code (1994)	Yes	3
New Orleans, LA	Yes	None	No	1
Las Vegas, NV	Yes	Adopted Model Energy Code	Yes	3
Long Beach, CA	Yes	Yes	No	2
Albuquerque, NM	Yes	Yes, included in 1994 strategic plan	Yes	3
Fresno, CA	No	Senior Citizen Weatherization	No	1
Virginia Beach, VA	No	None	No	0
Atlanta, GA	Yes	Yes	Yes	3
Mesa, AZ	Yes	None	No	1
Tulsa, OK	Yes	None	No	1
Miami, FL	No	Green Building Program	Yes	2

also determined that Sunbelt cities vary in their approaches to implementing energy-related policies. In addition, three specific locally adoptable energy-related policies were considered in detail: (1) city-operated energy efficiency programs, (2) local governmental energy program support, and (3) Energy Star^{TM} program participation. These policies are qualitative indications of the types of programs being pursued by the 25 sampled Sunbelt cities.

The specific energy-related policies adopted by each city were discussed in detail, and the sorts of policies in effect were identified. Evidence was provided that in many Sunbelt cities, policies are in effect to manage and reduce urban energy use. While these policies can be categorized as indicators of energy policy, local policy efforts, and organizational memberships, it is clear that there are variations in the themes of how these policies are locally defined and placed into practice.

Also notable are the examples of cities that use energy policy and energy conservation goals in their agendas as a means of achieving sustainability. Atlanta's program includes an internal energy conservation initiative. Tucson is developing a sustainable community based on use of solar energy. Mesa has created a planning agenda based on sustainability and is among those that have established an energy conservation program for city owned buildings.

It was discovered that the sampled Sunbelt cities vary broadly in their selection and application of policies. Certain cities, including Atlanta, Los Angeles, and San Diego, aggressively pursue multifaceted policies and focus their resources and agendas accordingly. On the other hand, most Sunbelt cities are more selective and limited in their policy choices. Cities such as Charlotte, Virginia Beach, and Oklahoma City are among those that tend not to adopt energy-related local policies.

Finally, it was determined that cities that have identified sustainability as a goal or have related sustainable development policies have substantially higher energy policy adoption rates than those that do not, when three select energy-related policies are used for comparative purposes. This suggests a positive relationship between the adoption of sustainability as a local governmental goal and the implementation of local energy related policies in Sunbelt cities. This research suggests that cities with sustainability as a local goal are more likely to adopt certain energy-related policies. Individuals and organizations seeking ways to get energy-related policies adopted by local governments in the Sunbelt might benefit by promoting sustainability as an urban goal.

REFERENCES

1. Bell, S.; Morse, S. *Sustainability Indicators, Measuring the Immeasurable*; Earthscan: London, 1999; pp. 103–124.
2. Beatley, T. *Green Urbanism: Learning from European Cities*; Island Press, Sage Publications: Washington, DC, 2000; 17.
3. Portney, K.E. *Taking Sustainable Cities Seriously*; The MIT Press: Cambridge, MA, 2003; 213.
4. City of Atlanta Online. *Energy conservation program history*, www.ci.atlanta.ga.us.mayor.energyconservation, 2003; 1.
5. City of San Antonio. *City of San Antonio Public Works*, www.sanantonio.gov, 2002.
6. City of Tucson. *Civano impact system memorandum of understanding in implementation and monitoring process*, June 26, 1998; 1.
7. Presidents Council on Sustainable Development. *Sustainable America—A New Consensus*; U.S. Government Printing Office: Washington, DC, 1996.
8. Roosa, S.A. A discussion of the economic aspects of implementing energy conservation opportunities. In *Solutions for Energy Security and Facilities Management Challenges*; Association of Energy Engineers: Atlanta, 2002; pp. 323–325.
9. United States Department of Energy. *Weatherization and intergovernmental program—rebuild America*, http://www.rebuild.org/index.asp, 2003.
10. City of San Diego. *City of San Diego Mayor Dick Murphy commissions the city's first "energy independent" municipal building utilizing solar power*, Environmental Services News Release, October 18, 2002; 1.
11. United States Environmental Protection Agency. *Join Energy Star—improve your energy efficiency*, www.energystar.gov, 2004.

Sustainable Building Simulation

Birol I. Kilkis
Green Energy Systems, International LLC, Vienna, Virginia, U.S.A.

Abstract
As the need for sustainable development increases, building simulation is becoming more crucial, and it is heading towards new challenges, dimensions, and concepts beyond the building envelope. Buildings are not isolated entities, which are not just responsible for about 35% of the total annual energy demand, but dynamically interact with their environment whose affected perimeter may be far wider than thought. In order to expand the building simulation perimeter to the entire impacted environment, all relevant variables must be factored in on a common base with a uniform metric. Among many attempts in this direction, exergy analysis establishes a uniform metric on all grounds and promises to make building simulation tools to more environmentally conscious.

INTRODUCTION

The objective of this entry is to highlight the importance of building simulation for analyzing and designing truly green and high performance buildings for a sustainable future. It also describes the need for new simulation tools that can cover a wider energy utilization and environment window, including the second law of thermodynamics, in a wide scope beyond the building envelope. More than ever, we need to know how efficient, stable, safe, functional, comfortable, and environmentally sustainable a building is under different indoor and outdoor conditions, functional and occupational requirements, and various equipment and system dynamics. Existing building simulation (BS) tools generally apply to a single building in a narrow window of the environment. These simulation and modeling tools allow us to design with methods to analyze, optimize, and control a building by calculating and compiling the crucial information regarding the overall performance of the building as accurately, precisely, and completely as possible, so that energy savings and a comfortable indoor environment are achieved in the building envelope with minimum possible cost. Only a few of the latest tools, however, include an environmental footprint analysis of a building, and these are rather limited.

NEED FOR BUILDING SIMULATION

People spend about two-thirds of their time indoors, with diverse and often conflicting needs relating to indoor air quality. According to several standards, including American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Standard 90.1 and ASHRAE Standard 90.2,[1,2] the primary targets are energy efficiency and indoor air quality, which conflict. According to the U.S. Department of Energy (DOE) Energy Efficiency and Renewable Energy Office (EERE), there are 81 million buildings in the U.S., of which 75% were built before 1979 and need a substantial retrofit or replacement of their Heating, Ventilating, and Air-conditioning (HVAC) systems. Fifteen million more buildings are expected by 2010. As a result, residential and commercial buildings in the U.S. are responsible for about 39% of the annual U.S. primary energy consumption, more than 70% of the total electric power consumed,[3] and close to 40% of CO_2 emissions (Fig. 1).[4] Buildings, which use natural gas for HVAC and domestic hot water, produce 20% of the total CO_2 gas emissions, estimated to be responsible for 60% of the greenhouse effect.

Consequently, buildings are the focal point of the sustainability quadrilemma of energy, economy, environment, and people. In order to resolve this quadrilemma, new building and retrofit designs must establish an optimum balance among all elements of environment, economy, people's needs, and rational use of energy resources. If it were possible to utilize the abundant, unused, low-exergy renewable and waste energy resources, all existing buildings could be heated and cooled. However, conventional HVAC systems cannot couple with low-exergy energy resources directly—because of their temperature incompatibility with low-exergy energy resources—unless HVAC equipment is oversized, or the resource temperature is conditioned. Both measures are cost-, energy-, and space-intensive, which greatly diminish the advantages of renewable and waste energy resources. This complexity requires developing new HVAC systems, with an outreach to the environment beyond the building envelope and a special emphasis on building-environment-energy source

Keywords: Building simulation; Exergy efficiency; Low-exergy buildings; Sustainable buildings; Sustainable development.

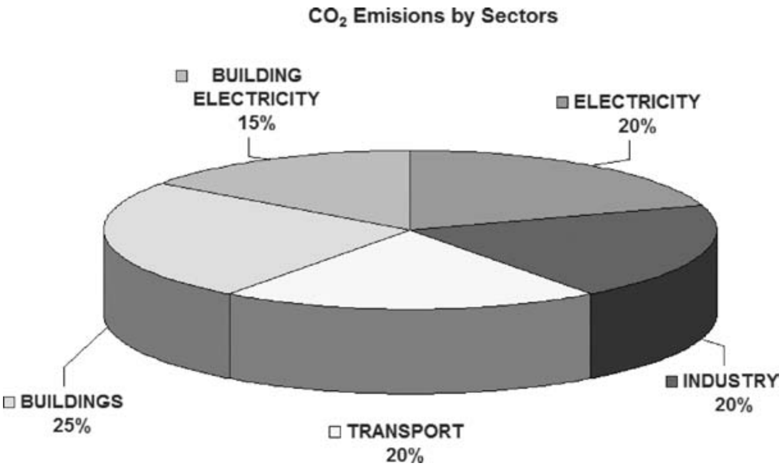


Fig. 1 Carbon emissions from fossil fuels in different sectors.
Source: From Refs 3 and 4.

relations. Today's green buildings may not be truly green unless the simulation window expands beyond the building envelope, with a clear understanding of exergy. Fig. 2 shows sample layers and scale of ideal building simulation windows for sustainability. Today there are almost 300 building simulation tools in 21 countries. However, in spite of this large availability and easy access, they do not yet address the overall picture; they are limited to the building envelope or its close vicinity.

ENERGY AND EXERGY ASPECTS OF BUILDINGS

"The absurdity of cutting butter with a chainsaw is immediately obvious to anyone."[5] On the same token, a

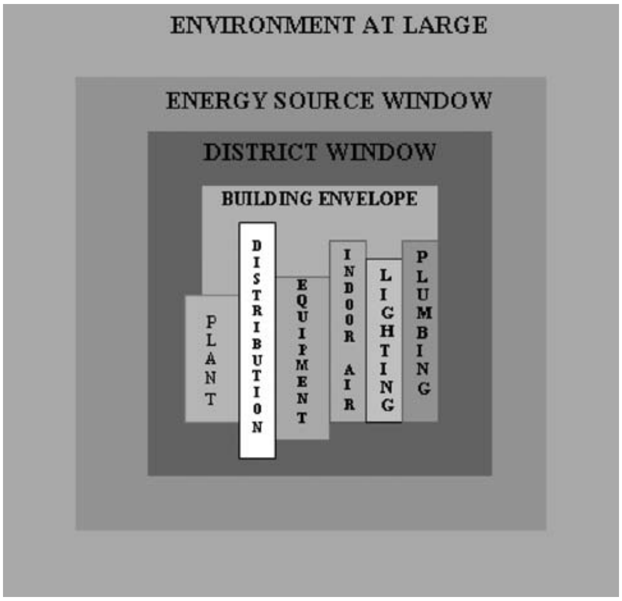


Fig. 2 Sample layers and scale of ideal building simulation windows.

conventional HVAC system uses high-grade energy in low-grade space heating or cooling and degrades the original energy resource, most often fossil fuel. According to Dincer and Rosen,[6] "Many scientists and engineers suggest that the impact of energy-resource utilization on the environment is best addressed by considering exergy. The exergy of a quantity of energy or a substance is a measure of the usefulness or quality of the energy or substance or a measure of its potential to cause change. In other words, exergy of any flow or resource of energy is the potential of useful work that is available, and a conventional HVAC system wastes most of it."[7] In fact, exergy is not only wasted but destroyed, because exergy flow is irreversible. Therefore, it is no surprise that the exergy efficiency of existing HVAC systems are less than 10%.[8] An earlier study showed that this efficiency is 5% on average for Swedish homes.[9] The shortcomings of the definition of energy efficiency are particularly apparent for tasks in which fossil fuels are used just for low-temperature heating. Since fossil fuels burn at very high flame temperatures—up to 2000 K[10]—the useful work potential (exergy) of fossil fuels is high. When fossil fuels are used for hot water heating, space heating, or even industrial steam production, most of the exergy is destroyed in these processes.[11] Indoor space heating furnaces have an estimated exergy efficiency of 6%, and heat pumps, when combined with conventional HVAC systems are not much better at 9%.[12] On the other hand, thermal efficiency of HVAC systems has almost reached a saturation point, well above 90% on average, except for thermal energy transport and distribution losses. Therefore, according to Annex 37,[13] the priority must now be shifted towards exergy efficiency in improving buildings, starting from their root causes. Fortunately, most of the root causes can be detected by developing next-generation building simulation codes that address the exergy efficiency at different scales—most importantly the HVAC system, which must be related to the environment and to the primary energy at its source, including the

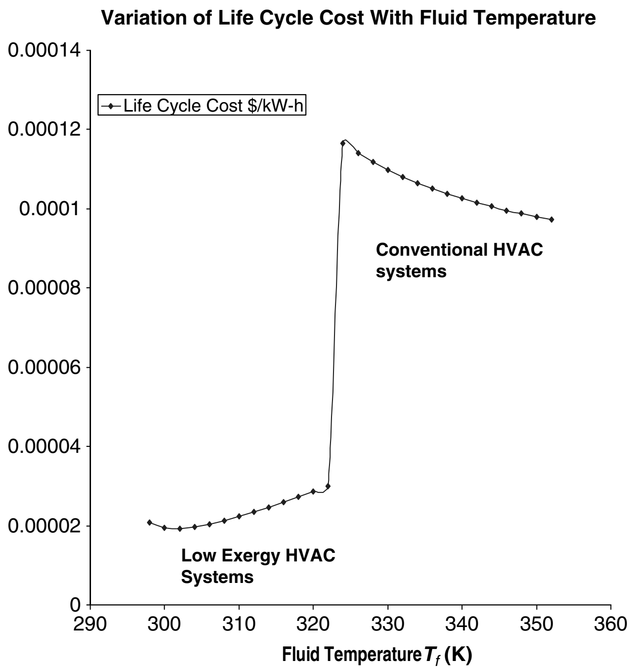


Fig. 3 Heating, ventilating, and air-conditioning life cycle cost including the cost of exergy destruction.
Source: From Ref. 16.

exergetic importance of thermal energy storage systems.[14,15] The rational exergy efficiency (ψ) is the ratio of the minimum exergy required by a given HVAC load to the actually available exergy of the energy source used in satisfying that load.[16]

$$\psi = \frac{\varepsilon_{\min}}{\varepsilon_{\text{act}}} \tag{1}$$

The minimum amount of exergy required to satisfy a unit heating load for an indoor space at a dry-bulb air temperature T_a, in reference to the temperature of the environment T_g[17]:

$$\varepsilon_{\min} = \left(1 - \frac{T_g}{T_a}\right) \qquad \{T_g \neq T_a\} \tag{2}$$

If the reference temperature of the environment is the ground temperature of 278 K (5°C) in winter, and the indoor air design temperature is 291 K (18°C), the ideal $\varepsilon_{\min}$ for an HVAC system directly utilizing ground heat is 0.0447. For a conventional HVAC system, the available exergy of the used energy source relates to the resource temperature and T_g. For a natural gas-fired furnace with a flame temperature of 1273 K, the available unit exergy is:

$$\varepsilon_{\text{act}} = \left(1 - \frac{278}{1273}\right) = 0.7816 \tag{3}$$

The exergy efficiency using Eq. 1 is then 0.057 (6%). This means most of the exergy is destroyed. If the same indoor space could be directly heated by an energy source at 302 K (29°C), Ψ could be much higher: 56%. These statements and equations are also valid for electric-based chiller or heat pump space cooling, if electric power is generated in a thermal plant. Fig. 3 illustrates the contrast between low-exergy HVAC systems and conventional HVAC systems as a function of operating fluid temperature.[16] The contrast is a result of including the cost of exergy destruction in the life cycle cost (LCC) analysis. Generally, low-exergy HVAC systems are a hybrid of radiant and convective heat transfer equipment, which can operate at moderate fluid temperatures.[18]

ILLUSTRATIVE EXAMPLES

Fig. 4A–D illustrate the importance of a wider building simulation window for sustainability.

1. *A Conventional HVAC System with a Central Power Plant*. In Fig. 4A, according to a conventional building energy simulation tool covering only the building envelope, the building seems to be energy efficient and may comply with most of the energy codes. Once a wider simulation window is used, covering the plant, the transmission lines, and the environment, it becomes apparent that the exergy efficiency is less than 10%.
2. *A Ground Source Heat Pump (GSHP) Building HVAC System with a Central Power Plant*. According to the same small building simulation window, the ground source heat pump in Fig. 4B makes sense only from the building owner/operator's side, because of a high COP. From a wider window, the overall efficiency remains almost the same and the exergy efficiency is still below 10%. The small window does not reveal how low the exergy efficiency is and does not show the ways and means to increase the exergy efficiency.
3. *A Decentralized Micro Combined Heat and Power (MCHP) Plant with a Conventional HVAC System*. In Fig. 4C, a wider simulation window shows that the overall energy efficiency is improving substantially, because power transmission losses are eliminated. Because the decentralized plant utilizes the primary energy resource for electric power generation on location, and captures most of the waste heat to provide thermal energy to the building, exergy efficiency has also improved. The exergy efficiency may further increase if a low-exergy HVAC system is installed, because waste heat can be more effectively utilized in the building.
4. *A Green Power (MCHP) System with Low-exergy HVAC System*. In Fig. 4D, both the energy and exergy efficiencies are high. Yet, this solution cannot be visible in a narrow simulation window.

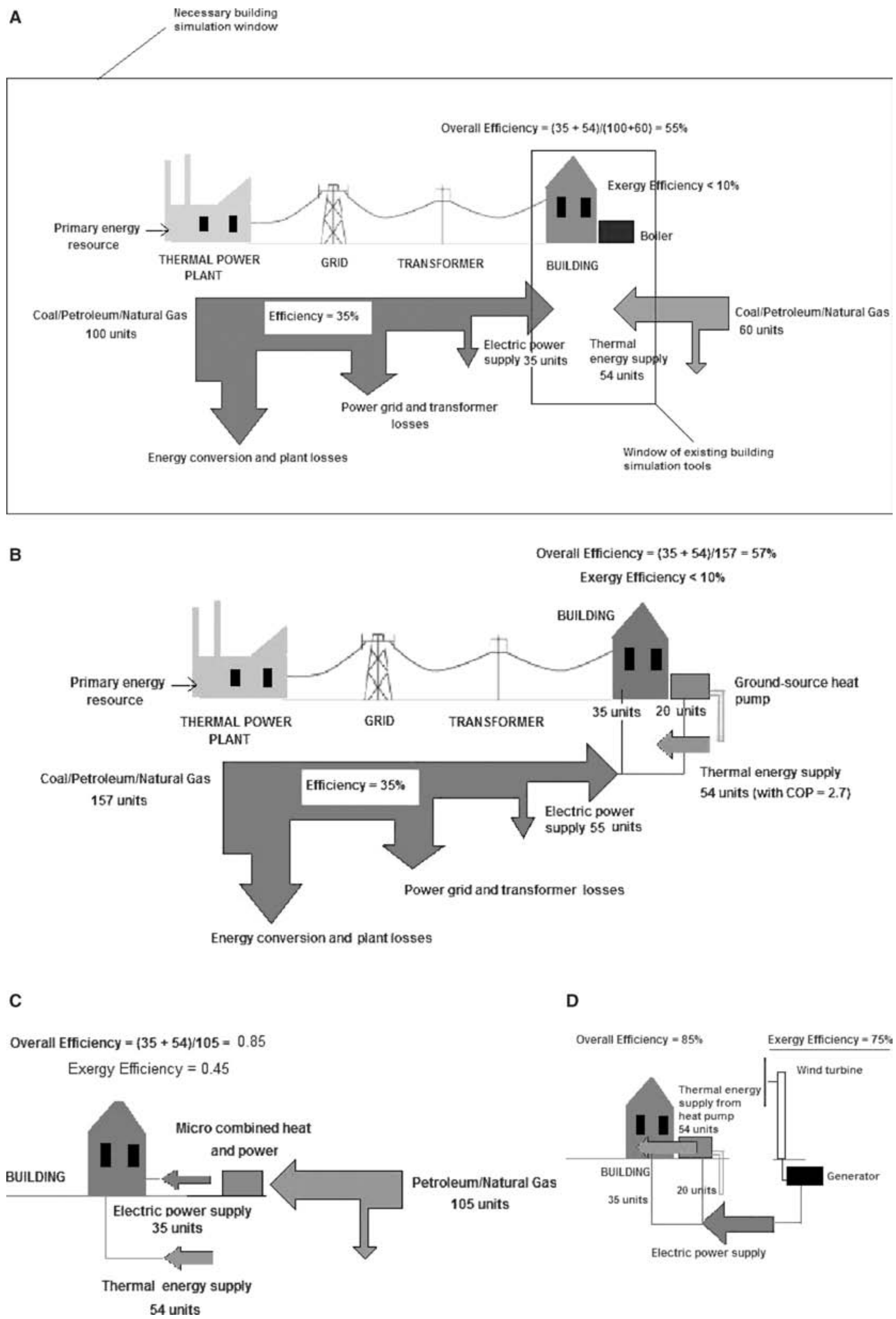


Fig. 4 (A) A centralized energy and power system with buildings using conventional heating, ventilating, and air-conditioning plant and equipment. (B) A centralized energy and power supply system with ground source heat pumps. (C) A decentralized micro combined heat and power system. (D) A decentralized, green energy system.

We can now conclude that whether it is thermal efficiency or exergy efficiency, or both, the simulation window must be far wider than the building envelope for sustainability. The building simulation window may only be reduced to building scale if decentralized energy systems with high efficiency green energy components are considered, as in Fig. 4D.[19]

MAJOR COMPONENTS OF BUILDING SIMULATION

Current theory of building energy simulation is based upon load and energy calculations developed by ASHRAE.[20] The selection of a simulation program for a given task depends on the project requirements, time, and cost of the analysis, experience of the user, and availability of suitable simulation tools and data.[20] Keeping in mind that future BS tools must include exergy analysis on a macro scale, we nevertheless currently need to select a BS tool best suited for a specific project. During the selection process, one needs to consider the following two requirements: (1) algorithms and data must be from reliable, well-established, published sources and (2) validation of the tool must be possible with available validation packages.[21] It is desirable to use a tool with an open structure or open end for future collaborative, modular versions. Expertise and training required must be compatible with the resources and background of the users. In addition, one must ask the following questions before the final decision:

- Who uses this tool? How many active users are there? Which institutions and countries use it?
- Is there a discussion group for this tool?
- What is the target audience? Does it suit your background, expertise, and position?
- What are the main features? Does it satisfy your needs?
- Is this tool compatible with other tools, programs, and databases? If yes, which programs are compatible? Can you bundle them?
- What are the required input and output data? How extensive are they, and what is their format and scope? Are there databases that can be readily used, or are databases already provided in the tool from recognized resources, and are they code approved?
- How many default inputs does the program tend to provide? Generally, more default inputs make it user-friendlier but substantially less accurate, depending on how the defaults are prepared and the assumptions involved.
- What is the computer platform, and which programming language is used?
- How much support can you get, and for how long can you get it? Can you get live support? Is the support free? How much and how detailed are educational materials, if provided?
- What are the costs, licensing options, terms, and conditions?
- Is a trial version available? What is the trial period and is it supported?
- Does the program check necessary code compliances? If yes, which codes?
- Will upgrades be available? What are the terms and conditions?
- What are the speed, capacity, and technical coverage of the program? Does it provide optimization and knowledge base tools?

The following basic components of building simulation must be covered:

- All sensible and latent loads
- Heat losses and gains from the envelope
- Internal gains and losses
- Electrical loads
- Water supply loads
- Waste management loads
- Ventilation loads for acceptable indoor air quality (IAQ)[22,23]
- Human comfort requirements and human behavior in controlled indoor environments
- Radiant asymmetry
- Mean radiant temperature
- Air velocity
- Wet-bulb and dry-bulb air temperature
- Relative humidity
- Mean radiant temperature
- Operative temperature
- AUST (Area Averaged Uncontrolled Surface Temperature)
- Convective and radiant heat transfer split
- Asymmetric thermal radiation
- Draft
- Vertical air temperature difference
- Warm or cold floors
- Hot ceiling
- Heat stress
- Comfort analysis[24,25]
- Duct losses[26]
- Domestic hot water demand modeling and supply[27]
- CBR (chemical, biological, and radiological) attack risk assessment and simulation[28]
- Fragility analysis (earthquakes, etc.)[29]
- Environmental ingress (like moisture penetration from foundations)
- Solar gains
- Shading
- Lighting simulation

INTEGRATED BUILDING DESIGN SYSTEM

Integration of simulation into the building design process can ensure that important data and information for each major design decision are provided in a timely fashion. By establishing design links and exchange between architecture and engineering, an integrated building design system (IBDS) can be developed.[30] Some researchers have taken the initiative to develop more efficient and flexible use of simulation tools. The COMBINE (Computer Models for the Building Industry in Europe) project[31] and the AEDOT (Advanced Energy Design and Operation Technologies) project in the U.S.[32] are typical examples. While design, simulation, operation, and control functions are becoming integrated, all equipment and systems must be bundled around a common protocol. Future integrated building simulation models must know exactly how each piece of equipment or sub-system behaves in the building through BACnet protocol,[33] and the equipment must be architectured accordingly.

AN OVERVIEW OF CURRENT BS TOOLS

According to G. Augenbroe,[34] "A broad range of simulation software applications has become available for a variety of building performance assessments over the last three decades …" The maturation of building simulation into a recognized and indispensable discipline for all professions—involved in the design, engineering, operation, and management of buildings—has now become the imminent challenge. Two key aspects dominate this evolution process: (1) attaining an increased level of quality assurance and (2) offering efficient integration of simulation expertise and tools in the overall building process. Major shortcomings of current simulation tools include:

- Input is usually lengthy and often cumbersome to prepare and compile.
- Parametric studies are not always possible.
- Outputs are generally too voluminous and need further analysis and interpretation.
- The learning curve may be too long and overly frustrating for the novice user.
- The user interface is generally overlooked and not user-friendly.
- The software is not flexible enough to test all possibilities.
- The software does not allow for the design and testing of new components. Off- the-shelf types of equipment must be used. This limitation reduces creative design opportunities.
- Most programs require long run times and a large memory. Personal Computer (PC) versions are available, but the compromises involved must be carefully weighed against the ease and simplicity of using them.
- Lifecycle analyses that include and recognize the exergy efficiency and the cost of exergy waste do not exist.
- Hybrid HVAC systems cannot easily be modeled.
- Optimization tools on a wide simulation window are not yet available.

Building simulation tools are available in the whole building level, equipment and component level, system level, retrofit level, and green building levels. Here only a very small number of available simulation tools are sampled. U.S. DOE EERE[35] provides a comprehensive and up-to-date listing and detailed information. Their Building Energy Software Tools directory lists almost 300 simulation programs categorized under: (1) whole-building analysis (load calculation, renewable energy, retrofit analysis, sustainability, and green buildings), (2) codes and standards, (3) materials, components, equipment, and systems, and (4) other applications.[36] A short list of typical BS tools is provided.

1. *APACHE*. Thermal design (heating, cooling, and latent load calculations). Equipment sizing, codes and standards checks, dynamic building thermal performance analysis, systems and controls performance, and energy use.[37]
2. *APACHE–HVAC*. Flexible and versatile system HVAC and controls modeling. Integrated simulation of building and HVAC systems.[38]
3. *BLAST*. The zone models of BLAST (Building Loads Analysis and System Thermodynamics) are based on the heat balance method. It performs hourly simulations of buildings, air handling systems, and central plant equipment. The output may be coupled to the LCCID (Life Cycle Cost in Design).[39]
4. *BSim2002*. Comprises different programs like a graphic model editor, thermal and moisture building simulation tool, dynamic solar and shadow simulation, daylight calculation tool, and compliance checker. Computer-aided design import and building integrated Photovoltaic system.[40]
5. *BEA*. Building Energy Analyzer is a system-screening tool to evaluate a variety of commercially available HVAC and power generation options. Uses the DOE-2.1E computational engine, includes a life cycle cost analysis module, and handles complex utility rates structures.[41]
6. *BUS*$^{++}$. New generation platform for building energy, ventilation, noise level, and indoor air quality simulations. A network assumption is adopted, and both steady state and dynamic simulations are possible.[42]

7. *DOE-2*. This is an hourly, whole-building energy analysis program that calculates energy performance and life cycle cost of operation. Can be used to analyze energy efficiency of given designs or efficiency of new technologies. Other uses include utility demand-side management and rebate programs, development and implementation of energy efficiency standards, and compliance certification. Training and expertise required.[43]
8. *EE4 CODE*. Used to determine the compliance of a building to Canada's Model National Energy Code for Buildings (MNECB). EE4 CODE may also be used to perform noncompliance energy analyses and thus to predict the annual energy consumption of a building and to assess the impact of design changes, based on DOE-2.1E.[44]
9. *EED*. A program for borehole heat exchanger design in a ground source heat pump system (GSHP) and borehole thermal storage. In very large and complex tasks, EED allows for the retrieval of the approximate required size and layout before initiating analyses that are more detailed.[45]
10. *EnergyPlus*. This is a whole-building energy simulation program that builds on BLAST and DOE-2. Includes advanced simulation capabilities, including time steps of less than an hour, modular systems simulation modules that are integrated with a heat balance-based zone simulation, and input and output data structures tailored to facilitate third-party interface development. EnergyPlus Version 1.2.2 was released in April 2005. EnergyPlus and weather data for more than 900 locations worldwide can be downloaded at no cost from the EERE home page.[46]
11. *EZDOE*. This tool is an easier-to-use personal computer version of DOE-2. EZDOE calculates the hourly energy use of a building and its life-cycle cost of operation, given information on the building's location, construction, operation, and heating and air conditioning system.[47]
12. *Right Suite-Residential*. All-in-one HVAC software performs residential loads calculations, duct sizing, energy analysis, equipment selection, cost comparison calculations, and geothermal loop design.[48]
13. *TRACE 700*. Follows the algorithms recommended by ASHRAE. Used for assessing the energy and economic impacts of building-related selections, such as architectural features, comfort-system design, HVAC equipment selections, operating schedules, and financial options.[49]
14. *TRNSYS*. TRNSYS (TRaNsient System Simulation Program) includes a graphical interface, a simulation engine, and a library of components that range from various building models to standard HVAC equipment, to renewable energy and emerging technologies. TRNSYS includes a method for creating new components that do not exist in the standard package.[50]
15. *VisualDOE*. A Windows interface to the DOE-2.1E energy simulation program. Users construct a model of the building's geometry using standard block shapes, using a built-in drawing tool, or importing DXF files. Building systems are defined through a point-and-click interface. A library of constructions, fenestrations, systems, and operating schedules is included, and the user can add custom elements. Up to 99 alternatives can be defined.[51]
16. *ESP-r*. Developed by the Energy Simulation Research Unit, University of Strathclyde, this is a general simulation tool that can be used to address a broad range of thermal performance problems, most often used for buildings.[52]
17. *HVAC Solution*. Allows users to graphically design and specify HVAC equipment, picking objects like boilers, pumps, fan coils, and air handlers. Using drag-and-drop methods, an HVAC system can be built. Once the system is built, the tool automatically renders equipment schedules and export schematics.[53]
18. *DUCTSIZE*. Calculates optimal duct sizes using the static regain, equal friction, or constant velocity method. Data entry can be accomplished manually or taken graphically from either Drawing Board or AutoCAD. A library of fan data for noise calculations is built into the program.[54]
19. *Hydronics Design Studio*. Assists in analyzing the thermal and hydraulic performance of modern hydronic heating systems in residential and light commercial buildings. The professional version performs tasks like heating load analysis, series baseboard circuit analysis, piping heat loss estimating, expansion tank sizing, radiant circuit analysis, injection mixing simulation, buffer tank simulation, and fuel cost comparisons.[55]
20. *FLOVENT*. Calculates airflow, heat transfer, and contamination distribution for built environments. FLOVENT uses techniques of Computational Fluid Dynamics (CFD) packaged in a form that addresses the needs of mechanical engineers involved in the design and optimization of ventilation systems.[56]
21. *ArchiPhysics-Solar*. A passive solar energy system for free running buildings. Shows the adaptive comfort level, as well as indoor and outdoor air temperatures for given building geometry, location, and glazing.[57]

22. *MOIST*. This program predicts the combined transfer of heat and moisture in multi-layer building construction. Inputs hourly weather data and predicts the moisture content and temperature of the construction layers as a function of time of year. It can be used to develop guidelines and practices for controlling moisture in walls, flat roofs, and cathedral ceilings.[58]
23. *Daylight*. This program calculates the daylight factor distribution in a room. A user-friendly program by Archiphysics.[59]
24. *BREEZE*. This is a tool for estimating ventilation rates, developed by Building Research Establishment (BRE) in England.[60]

SUSTAINABLE DESIGN AND GREEN BUILDINGS

There are currently some desktop tools available for architects and engineers for sustainable design of buildings. However, these tools perform life cycle environmental impact assessment based solely upon building materials and related items.[61–63] The energy balance, exergy balance, and thermo-physical interactions of the building with the environment are not included.

1. *ATHENA*. This is an environmental impact estimator program developed by ATHENA Sustainable Materials Institute. The program has a large database of building materials and performs an environmental impact assessment.[64]
2. *BEES*. The BEES (Building for Environmental and Economic Sustainability) software enables the use of cost-effective, environmentally preferable building products. The tool is based on consensus standards and includes actual environmental and economic performance data for nearly 200 building products.[65]
3. *RETScreen*. Developed by the Natural Resources of Canada, the RETScreen International Clean Energy Project Analysis Software is a decision support tool that can be used worldwide to evaluate the energy production, life-cycle costs, and greenhouse gas emission reductions for various types of energy efficient and renewable energy technologies. Also includes product, cost, and weather databases, as well as a detailed online user manual.[66]
4. *GBS*. Seamlessly links architectural 3-D CAD building designs with energy analysis. Green Building Studio (GBS) enables architects to quickly calculate the operational and energy implications of early design decisions. GBS uses the DOE-2 simulation engine to calculate energy performance and generates geometrical input files.[67]

More information can be found at the Building Energy Simulation Tools (BEST) Web site,[68] which lists the available tools and their main features and provides Internet sites and references. The most up-to-date information about the energy design tools can be obtained from the International Building Performance Simulation Association.[69]

CONCLUSION

With the ever-increasing need for high performance buildings with truly green features, it is virtually inevitable that future building simulation tools will become an integral part of other macro-scale simulation tools for the environment, the energy sector, and the community. In achieving this goal, building simulation tools must be more compatible, open-structured, and open-ended for greater modularity, compatibility, and data exchange. Finally and most importantly, we must include exergy-based optimization in the broadest simulation window possible within these future tools.

REFERENCES

1. ASHRAE *Standard 90.1-2004—Energy Standard for Buildings Except Low-Rise Buildings, SI Edition (ANSI approved; IESNA Co-sponsored)*; ASHRAE: Atlanta, GA, 2004.
2. ASHRAE *Standard 90.2-2004—Energy-Efficient Design of Low-Rise Residential Buildings (ANSI approved)*; ASHRAE: Atlanta, GA, 2004.
3. Torcellini, P.A.; Judkoff, R.; Crawley, D.B. High-performance buildings. ASHRAE J. **2004**, *46* (9), S4–S12.
4. DOE, U.S. Department of Energy. *Emissions of greenhouse gases in the United States-2 carbon dioxide emissions*, Report No. EIA/DOE-0573, 1998.
5. Simpson, M.; Kay, J. Availability, exergy, the second law and all that, http://www.fes.uwaterloo.ca/u/jjkay/pubs/exergy/, copyrighted in 1989.
6. Dincer, I.; Rosen, M.A. A worldwide perspective on energy, environment and sustainable development. Int. J. Energy Res. **1998**, *22*, 1305–1321.
7. Rosen, M.A.; Dincer, I. Energy and exergy analyses of sectoral energy utilization: an application for Turkey. In *Proceedings of first International Trabzon Energy and Environment Symposium*, Karadeniz Technical University; Tarbzon Turkey,1996; Vol. 3, 1059–1065.
8. Kilkis, I.B. An exergy aware optimization and control algorithm for sustainable buildings. IJGE **2004**, *1* (1), 65–77.
9. Wall, G. Exergy—a useful concept. *Chalmers Biblioteks Tryckeri*; Göteborg: Sweden, 1986.
10. Dincer, I.; Cengel, Y. Energy, entropy and exergy concepts and their roles in thermal engineering. Entropy **2001**, *3*, 116–149.
11. Ralph, T. *Half Life: Nuclear Power and Future Society*; Ontario Coalition for Nuclear Responsibility: Ottawa, 1981; 178.

12. *American Institute of Physics Efficient Use of Energy*; American Institute of Physics: New York, 1975; 49–50.
13. IEA and ECBCS *Guidebook to IEA ECBCS Annex 37—Low Exergy Systems for Heating and Cooling of Buildings*; VTT: Finland, 2003.
14. Dincer, I.; Rosen, M.A. *Thermal Energy Storage, Systems and Applications*; Wiley: West Sussex, 2002.
15. Dincer, I.; Rosen, M.A. Energetic, environmental and economic aspects of thermal energy storage systems for cooling capacity. Appl. Thermal Eng. **2001**, *21*, 1105–1117.
16. Kilkis, I.B. A road map for emerging low-exergy HVAC systems. Int. J. Exergy **2005**, *2* (4), 348–365.
17. Dincer, I.; Rosen, M.A. Thermodynamic aspects of renewables and sustainable development. Renew. Sust. Energy Rev. **2005**, *9*, 169–189.
18. Kilkis, I.B.; Suntur, A.S.R.; Sapci, M. Hybrid HVAC systems'. ASHRAE J. **1995**, *37* (12), 23–28.
19. Kilkis, S. *Development of a Rational Exergy Management Model to Reduce CO_2 Emissions with Global Exergy Matches,* Honors Thesis, Georgetown University, 2007.
20. ASHRAE *ASHRAE Handbook—HVAC Applications*; ASHRAE: Atlanta, GA, 2003.
21. ASHRAE *Standard 140-2004—Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs (ANSI Approved)*; ASHRAE: Atlanta, GA, 2004.
22. ASHRAE *Standard 62.1-2004—Ventilation for Acceptable Indoor Air Quality (ANSI Approved)*; ASHRAE: Atlanta, GA, 2004.
23. ASHRAE *Standard 62.2-2004—Ventilation and Acceptable Indoor Air Quality in Low-Rise Residential Buildings (ANSI Approved)*; ASHRAE: Atlanta, GA, 2004.
24. ASHRAE *Standard 55-2004—Thermal Environmental Conditions for Human Occupancy (ANSI Approved)*; ASHRAE: Atlanta, GA, 2004.
25. ASHRAE *ASHRAE Handbook—HVAC Systems and Equipment*; ASHRAE: Atlanta, GA, 2004.
26. ASHRAE *Standard 152-2004—Method of Test for Determining the Design and Seasonal Efficiencies of Residential Thermal Distribution Systems (ANSI Approved)*; ASHRAE: Atlanta, GA, 2004.
27. Tiller, D.K.; Phil, D.; Henze, G.P. *Online domestic hot water end-use database*, ASHRAE research project 1172 report; ASHRAE: Atlanta, GA, 2004.
28. Arvelo, J.; Brandt, A.; Roger, R.P. *An Enhanced Multizone Model and its Application to Optimum Placement of CBW Sensors*; ASHRAE: Atlanta, GA, 2002.
29. Wen, Y.K.; Ellingwood, B.R. The Role of Fragility Assessment in Consequence-Based Engineering, Earthquake Spectra, EERI, August, 2005.
30. Lewis, M. Integrated design for sustainable buildings. Building for the future: a supplement to. ASHRAE J. **2004**, *46* (9), 22–30.
31. COMBINE. http://erg.ucd.ie/combine.html.
32. AEDOT. http://apc.pnl.gov:2080/0projects_and_capabilities/aedot/html/aedot.html.
33. ASHRAE *Standard 135-2004—BACnet®—A Data Communication Protocol for Building Automation and Control Networks (ANSI Approved)*; ASHRAE: Atlanta, GA, 2004.
34. Augenbroe, G. Trends in building simulation. Building Environ. **2002**, *37*, 891–902.
35. U.S. DOE EERE. www.eere.energy.gov.
36. http://www.eere.energy.gov/buildings/tools_directory/subjects.cfm/pagename = subjects/pagename_menu = whole_building_analysis/pagename_submenu = energysimulation.
37. APACHE. Thermal Design (Heating, Cooling & Latent Load Calculations) http://www.ies4d.com.
38. APACHE–HVAC. http://www.ies4d.com.
39. BLAST. http://www.bso.uiuc.edu.
40. BSim2002. http://www.bsim.dk.
41. BEA. http://www.interenergysoftware.com.
42. BUS++. Pekka.Tuomaala@vtt.fi.
43. DOE-2. http://simulationresearch.lbl.gov.
44. EE4 CODE. http://ee4.com.
45. EED. http://www.buildingphysics.com.
46. EnergyPlus. http://www.eere.energy.gov/buildings/energyplus/ or EnergyPlus Web site.http://www.energyplus.gov.
47. EASYDOE. http://www.elitesoft.com.
48. Right Suite-Residential, http://www.wrightsoft.com.
49. TRACE 700. http://www.trane.com/commercial/software.
50. TRSYS. http://sel.me.wisc.edu/trnsys/downloads/download.htm.
51. VisualDOE. http://www.archenergy.com/products/visualdoe/.
52. ESP-r. http://www.strath.ac.uk/Departments/ESRU/Ad_Env_Sim.html.
53. HVAC Solution. http://www.hvacsolution.com.
54. DUCTSIZE. http://www.elitesoft.com.
55. Hydronics Design Studio. http://www.hydronicpros.com.
56. FLOVENT. http://www.flovent.com.
57. ArchiPhysics-Solar. http://www.archiphysics.com/programs/freerunner/freerunnner.htm.
58. MOIST. http://www.bfrl.nist.gov/863/moist.html.
59. Daylight. http://www.archiphysics.com/programs/daylight/daylight.htm.
60. BREEZE. http://www.bre.co.uk/.
61. Gowri, K. Desktop tools for sustainable design. ASHRAE J. **2005**, *47* (1), 42–46.
62. Grumann, L.D., Ed., *ASHRAE GreenGuide*, ASHRAE: Atlanta, GA, 2003; 190.
63. GBC. LEED Reference guide, Version 2.0, U.S. Green Building Council, 2001.
64. ATHENA. www.athenasmi.ca.
65. BEES. http://www.bfrl.nist.gov/oae/software/bees.html.
66. RETScreen. http://www.retscreen.net/.
67. Green Building Studio GBS. http://www.greenbuildingstudio.com/.
68. Building Energy Simulation Tools (BEST). http://www.arch.hku.hk/research/BEER/best.htm#Simulation%20Methods.
69. International Building Performance Simulation Association. http://next1.mae.okstate.edu:80/ibpsa/.

BIBLIOGRAPHY

Additional On-Line Resources

1. *IBPSA*. The International Building Performance Simulation Association (IBPSA), (http://www.ibpsa.org) is a nonprofit international society of building performance simulation researchers, developers, and practitioners, dedicated to

Solid–Sus

improving the built environment. Since 1985, IBPSA has been organizing international conferences on building simulation every two years. PDF versions of all papers from the 1985–2003 Conferences are available online. http://www.ibpsa.org/m_events.asp.

2. *BLDG–SIM*. This is a mailing list for users of building energy simulation programs. This allows engineers, architects, and others in the building design trade to compare alternative designs and select the design that is cost justified. The Web page for this mailing list is located at http://www.gard.com/ml/bldg-sim.htm.
3. *Center for Buildings and Systems-TNO-TU/e*. The institute's multidisciplinary approach is expected to result in innovative insights in the field of lighting, indoor climate control, noise, health, and comfort, but also with regard to sustainable energy. Their studies include building simulation, acoustics, lighting, renewable energy systems, health, comfort, and productivity. http://www.kcbs.nl.
4. *IEA ECBCS*. The mission of Energy Conservation in Buildings and Community Systems is to facilitate and accelerate the introduction of energy conservation and environmentally sustainable technologies into healthy buildings and community systems, through innovation and research in decision-making, building products and systems, and commercialization. http://www.ecbcs.org.
5. *LowEX (Low Exergy Systems for Heating and Cooling of Buildings)*. The LowExNet has been founded because of the work of the International Energy Agency (IEA) Annex 37: "Low Exergy Systems for Heating and Cooling in Buildings." During an expert meeting in Kassel, Germany in September 2003, the decision was made to form a network on the issues of new ways for energy systems in buildings and to continue the work of the IEA Annex 37, which is to seek various low exergy system solutions. http://www.lowex.net/english/inside/links.html.
6. *ZUM (Zentrum für Umweltbewusstes Bauen e.V. Center for Sustainable Building)*. This institution provides several downloadable programs for building simulation with emphasis on sustainability. www.zub-kassel.de.

Ashrae Guides

1. Advanced Energy Design Guide for Small Office Buildings.
2. Humidity Control Design Guide for Commercial and Institutional Buildings.
3. Practical Guide to Noise and Vibration Control for HVAC Systems (SI).
4. Application Guide: Indoor Air Quality Standards of Performance.

Selected Books

1. Clark, J.A. *Energy Simulation in Building Design*, 2nd Ed.; Heinemann: Butterworth, 2001.
2. Waltz, J.P. *Computerized Building Energy Simulation Handbook*; Fairmont Press: Lilburn, Georgia, 2000.
3. Taplin, H. *Boiler Plant and Distribution System Optimization Manual*, 2nd Ed.; Fairmont Press: Lilburn, Georgia, 1997.
4. Curl, R.S. *Building Owner's & Manager's Guide: Optimizing Facility Performance*; Fairmont Press: Lilburn, Georgia, 1998.
5. Petchers, N. *Combined Heating, Cooling & Power Handbook: Technologies & Applications*; Fairmont Press: Lilburn, Georgia, 2002.

Sustainable Development

Mark A. Peterson
Sustainable Success LLC, Clementon, New Jersey, U.S.A.

Abstract
Sustainable development includes all business and community planning and operating decisions with due consideration for: (1) people—employees, customers, shareholders, community residents, or anyone that is involved or affected; (2) planet—material and energy resource management that does not hurt the environment; and (3) profits—or economics or prosperity. Sustainable development takes a different, more caring look at how people interact with themselves and how their activities affect the planet and the general well being of life for sustained economic growth.

INTRODUCTION

This article defines sustainable development and its three basic aspects. Because sustainable development is a relatively new concept, a short history and description of the drivers that lead to sustainable development are described. Then, a sustainable energy future is presented. To be sustainable as a society requires cooperation and collaboration rather than command and control management. A very key aspect of sustainable development, social synergy, is covered. Also, because sustainable development is relatively new but essential to build a better and viable future, children from the earliest age through college need to learn to understand and apply the concepts of sustainable development. A paragraph is included that describes current efforts in the United States to incorporate education in sustainable development (ESD) in K-12 and college curricula.

Sustainable development encompasses stewardship of many areas of human and planetary life. In business, one of the key motivators is to implement sustainable development measures to be profitably successful indefinitely. To do this requires that businesses show their due diligence to both society and the environment while maximizing profits. Sustainability reporting assists businesses in assessing their efforts. Sustainability reports are both management and public relations tools. An innovative, very effective, time-tested form of sustainable development, "renting a service" rather than "selling a product," is covered. Then, sustainable development for community vitality is described. As the sizes, types, and socioeconomics of communities vary considerably, so do all of the related aspects of evolving them to be sustainable. Reference is made to a web site that thoroughly describes all aspects. Some of the subjects covered on the web site are briefly summarized in this article. The final section covers the vast, opening market of sustainable development in developing countries.

Keywords: Sustainable; Sustainability; Development; Triple bottom line; World; Business; Communities; Success; Social; Responsibility.

WHAT IS SUSTAINABLE DEVELOPMENT?

Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their needs. Sustainable development has three aspects:

1. Social (people).
2. Environmental (planet).
3. Economic (profits)/prosperity.

All development affects all three aspects. All three aspects are interdependent. Thus, being mindful of these interdependencies in management and leadership decisions will result in the best overall solution—a win–win–win solution that maximizes success and minimizes any negative social, environmental, and economic costs. This is called managing the triple bottom line of people, planet, and profits. This is also called whole systems thinking[1,2] because all relevant factors are considered as a whole. The role of engineers is to help their clients be successful. This requires integrated whole systems thinking that covers all related liabilities that a company or community (their client) may have and provides the most efficient and profitable solution to the challenge. Often, the best whole system solution is also the most efficient and most sustainably profitable.

The environmental (planet) aspect is significantly affected by energy consumption and management, including: the entire national power infrastructure and distribution, transportation, plus the construction and renovation of all residential, commercial, and industrial facilities.

HISTORY, ENVIRONMENTAL DEGRADATION, AND NATIONAL SECURITY

During the last century, while fossil fuels were abundant and cheap, those fuels fulfilled a majority of our energy conversion needs. The mounting problem is that combustion emissions have fouled the environment in a number of ways, resulting in increases in respiratory illnesses, mercury pollution, and a rise in global temperatures. The quantity of easily retrieved fossil fuels is significantly depleted. Coal is still relatively abundant, but it does not burn cleanly. Technologies need to be developed to both mine the coal safely and to burn it cleanly. Regarding petroleum, many countries that are not friendly to the United States control most of the remaining easily extractable sources. Many national security advisors have indicated the urgency of severing our dependence on foreign oil as a part of an overall strategy for the security of the United States and as a means to prevent oil-related conflicts.[3,4] It is also apparent that there is a need to protect and allow the environment to regenerate. The effect of using fossil fuels extensively and inefficiently is that we are simultaneously poisoning the environment and ourselves. Nuclear power emissions are clean, but the nuclear power industry has significant obstacles such as storage of radioactive wastes for many thousands of years. In addition, there are security concerns to safeguard radioactive material from being stolen for production of atomic weapons.

Governors and mayors are taking action to implement clean energy technologies. In response to clear signs of increased cost from continuing to use fossil fuels and scientific evidence showing that by burning fossil fuels we are initiating a possibly devastating global warming trend[5] that could flood coastal cities, disrupt the food chain, and change climate patterns significantly, many states have taken the initiative and enacted renewable energy portfolios to fund the transition to renewable energy resources. Many remaining states are in the process of developing their own renewable energy portfolios. These renewable energy portfolios provide significant state- and utility-sponsored financial incentives for the commercial, industrial, and residential use of renewable energy systems and fuels. On the city level, many mayors from major cities around the world have made commitments to cut greenhouse gas emissions to slow the rate of global warming.[6] Many of these cities are coastal and could be severely impaired or destroyed from rising sea levels from global warming. So, civic action to switch to cleaner energy options is beginning in earnest.

THE ENERGY FUTURE

What this means to energy engineers is that petroleum-derived fuels are on their way out over the next half century. Hydrogen (where the hydrogen is derived from renewable energy sources), ethanol, biodiesel, and other forms of renewable fuels are on their way in. Direct and indirect conversion of solar energy, including wind, biomass, wave/tidal power, and small-scale hydroelectric power will increasingly be part of the energy infrastructure that energy engineers will design and build. The bottom line with energy is that it needs to be relatively nonpolluting and indefinitely available. It is a very dynamic time for energy engineers as the entire, world-wide energy picture transitions to clean renewable technologies. This will eventually add a lot of stability to the world economy, the world political environment and to everyone's lives. The stability will come from the fact that renewable energy technologies can be used to tap the natural energy resources that are available everywhere. Stability will also come, as the environment regenerates, the climate stabilizes and resources remain available for our sustenance.

HOW WE SOCIALLY AND PROFESSIONALLY INTERACT WITH EACH OTHER DETERMINES OUR DEGREE OF SUCCESS

Another aspect of sustainability has to do with how well we interact and collaborate. In the past, management of most activities was by a top-down hierarchy. Now humanity is evolving and it is driven by high levels of sophistication in technology and communications, which has resulted in individual knowledge and skill level increases. Plus, many families are now structured such that individual adults and children are taking more independent responsibilities for the many aspects of their lives. This has all created a desire by many people to be more intimately involved in solutions rather than just letting someone else do the thinking. In this new economy, teamwork and open communication are important to bring all stakeholders together, whether in a business or community, to manage by consensus and cooperation. Everyone affected should have the opportunity to be involved in the solution, even if just by being informed as the planning and decision-making are in process. The outcome then is one that promotes efficiency, for the simple reason that all persons affected are involved, which promotes enthusiasm and "buy-in." Typically with this process, more work and time are invested up-front such that all aspects are considered and thus everything proceeds more efficiently down-stream.[7]

CREATIVE, COOPERATIVE, DESIGN AND PLANNING TEAMWORK[2]

The ASHRAE GreenGuide recommends "integrated design teams" that have all of the design, economic, planning, and other related disciplines involved up-front to create better designs. If this process is not used, design typically proceeds in a series of "handoffs" that tend to compound problems, as each succeeding team designs "around" any incompatibilities that the previous designers have already finished. This adds unnecessary complexities and inefficiencies, which increase construction and life-cycle costs. Through coordination and the collaboration of designers, architects, engineers, and key players, an integral design can be created that functions as an efficient system and not as a collection of parts that are force-fit together. This approach has enabled design teams to design very energy efficient, comfortable, aesthetically and environmentally friendly buildings at or less than the conventional price per square foot of traditionally designed buildings. So, working as a team from the beginning is the most efficient way of designing because potential conflicts are resolved up-front rather than later at a higher cost. Effectually, when people are creating and they know that their contributions are respected, superior planning and designs are achieved. The upfront work of coordinating planning and brainstorming sessions with many people of diverse backgrounds can be a challenge. However, the results and life-cycle costs are almost assuredly optimal.

EDUCATION

The United Nations has declared the decade from 2005 to 2014 as The Decade of Education for Sustainable Development.[8]

There are many national teams around the world that have taken the lead to work across the educational spectrum—including public and private education, primary and higher education, independent, charter, and home-schooling—to incorporate ESD in their curricula. The U.S. Partnership for The Decade of Education for Sustainable Development[9] was formed to facilitate implementation of ESD in the United States.

Effective education can demonstrate the inter-relationship and interdependence of people, planet, and profits in all life activities. We have the technology to transition to a clean energy future and to manage materials in a cyclic manner. We know that pollution is causing global environmental change. We also know that teamwork and cooperation get better results than working in a hierarchical or isolated manner. Education in sustainable development will show that the current "consume and throw away" economy no longer works for the benefit of humanity and life. Rather, there is a need for cyclic, whole-systems thinking that integrates all relevant factors into the best, longest lasting results.

SUSTAINABILITY REPORTING

Realizing the importance of being "sustainable" and understanding that they are good long-term investments, many companies are developing sustainability reports. These reports are strong management tools that show how well a company is progressing with their sustainability, their corporate social responsibility (CSR), and their goals to continually improve. They provide an openness that stakeholders (employees, stockholders, regulatory agencies, customers, and community leaders) expect so they know if a company is a good place to work, a good investment, or a good neighbor. These reports are available to the public (free download).[10] For engineers, preparation of sustainability reports involves the collection and analysis of energy and environmental data. Demonstration of energy savings plans and associated pollutant emission reductions can have significant public relations and market share value. This can promote higher sales and flow of investor capital as companies prove their social responsibility and long-term viability.

"High 5!—Communicating your Business Success through Sustainability Reporting—A Guide for Small and Not-So-Small Businesses," from the Global Reporting Initiative, describes the benefits of sustainability reporting:

"Sustainability reporting has many advantages that benefit different areas of your business. Some benefits are purely financial while others deal with customer or employee satisfaction. Sustainability reporting helps organizations identify and address their current and potential risks, saving time and money in the short and long term. As the public becomes more aware of your efforts, customer loyalty and credibility of your business will greatly increase. When you take a deeper look into your daily business operations through sustainability reporting you will be able to discover new opportunities.

Businesses continuously seek to generate income and acquire a competitive edge by identifying new market opportunities and determining current and potential risks. When your organization embraces a sustainability perspective, i.e., simultaneously addressing social, environmental and economic issues, you can benefit from cost-savings and improvements in product quality and employee performance.

When considering how to improve financial performance, many organizations only look at financial aspects, such as the cost of purchasing goods, personnel costs, or tax payments. However, working on environmental and social issues can also positively affect your financial bottom line.

Sustainability reporting is the way to identify these potential benefits and realize the economic gains. It helps you achieve your business goals by setting up a continuous improvement process based on target setting and progress measurement. All in all, sustainability reporting helps you to acquire that competitive advantage."

RENTING VS BUYING (A SUSTAINABILITY INNOVATION)[11,12]

Here is an example of systems thinking to ensure that a product has minimal environmental impact plus high social and economic value:

In today's consumer/throw-away economy, typically a product is manufactured and sold. There is producer incentive to minimize the use of labor and material resources put into a product, thus saving on cost. The product is sold as cheaply as possible to maximize sales. The product eventually wears out and is disposed of. A product that breaks and is disposed of soon after its warranty period is best for sales, so that a consumer will go out and buy a replacement. This is obviously a wasteful scenario, which is prevalent in commerce today. However, that is changing.

A more efficient scenario that many companies have successfully deployed for years is to manufacture and rent a product. Then, there is an incentive to maximize the utility and life of the product, thus ensuring an income to the owner/renter for as long as possible into the future. (This is what sustainable development is all about—maintaining economic flow for as long as possible into the future.) When this product is manufactured, due care in manufacturing processes and sufficient material are used to maximize durability, reliability, and longevity. Customer satisfaction is high from having a reliable product, which also builds name recognition and increases desirability. The product is designed to be easily maintainable, again, to maximize longevity, and also to facilitate dismantlement of the product when it has come to the end of its useful life. Thus, the parts can be easily remanufactured and reused or segregated for efficient recycling of the raw materials. So, maximum utility and income is achieved from the product, and most of the resources that went into manufacturing and maintaining the product throughout its life are recycled with minimal impact to the environment.

So, engineers with sustainable development in mind might be thinking of the "rented" product scenario, which is inherently efficient, rather than the "sold" product scenario, which is inherently wasteful. For example, Interface, Inc., the largest carpet manufacturer in the world and a company that has committed itself to be as sustainable as possible, is using this renting concept in its products. They rent carpet tiles. As the carpet eventually wears, tiles are replaced and recycled to make new tiles.

COMMUNITY DEVELOPMENT TO DECREASE ENERGY COSTS[13]

Over the last three-quarters of a century, there was an assumption that gasoline and diesel fuel would be cheap and plentiful, indefinitely. Thus, urban sprawl developed because of the low cost of owning and operating one or more cars. This is not the case anymore! Now, with the realization that cars are expensive to both purchase and operate, there are efforts by many cities to re-establish neighborhoods that have all of the amenities needed for occupants all within a short distance that can be covered on foot, with a bicycle, or with convenient public transportation. These cities are excluding automobiles from certain areas and some are charging admission fees for cars to enter semirestricted areas. This has helped to revitalize many city commercial districts because of the park-like feeling of being in these areas without the noise and exhaust from cars. With the economic benefits that have been realized by these arrangements, more and more urban and suburban cities are pursuing these types of commercial district renovations in their cities.

Obviously, society has made a major investment that established urban/suburban sprawl. That investment now needs to be made sustainable. Consequently, there will be major investments in clean fuels such as biodiesel, ethanol, and renewably derived hydrogen to power the huge fleet of vehicles in the United States. A remaining economic burden to maintain suburban living is the maintenance of roads. However, that may be relieved as convenient, modern, and cleaner public transit busses, trains, and other guided vehicles are developed into transportation networks, thus decreasing the number of cars on the road. Energy engineers will be integral to this transition in transportation.

So, the paradigm for energy engineers will evolve as community structures around urban and suburban cities change and improve for energy efficiency. There will probably be a tendency to create nodes, where residents have a majority of the amenities nearby. Then, clean, comfortable, energy-efficient public transportation and express lanes for cars will connect each node.

OVERVIEW OF THE MANY ASPECTS OF SUSTAINABLE COMMUNITIES INFRASTRUCTURE AND NATURE

The web site http://www.conservationeconomy.net/INDEX.CFM (courtesy of the Ecotrust) summarizes the many and various aspects of sustainable communities and preserving our natural assets. Many social factors that play a role in the culture and systems changes associated with sustainable development are thoroughly explained on the web site. Engineers should know these social interactions, which are cohesive and essential to sustainable communities. Readers are encouraged to visit the web site to get a flavor of all of the aspects of sustainability or to narrow in on aspects that are of particular interest to them. The following is a narrative summary of some of the chapters on the web site:

A Conservation Economy describes how social capital, natural capital, and economic capital can be synergetic and sustainable.

Social Capital covers fundamental needs, which include: the strong need for local sources of food; accessible, healthy shelter; healthy environment and access to healthcare; plus access to knowledge about the interconnectedness of us to our environment and to each other. The section on community discusses collaborative processes that honor: social equity, which promotes prosperity for all; security from fear and violence; recognition of the wealth and strength in our cultural diversity and establishment of a will to preserve it; plus, local celebrations to honor a sense of place with the community and environment. It proceeds to describe the importance of enjoying beauty and play, to relieve stress from our busyness; learning to welcome transitions that improve communities as a whole; and establishment of civic society where all residents can manage their communities collaboratively.

Natural Capital includes the atmosphere, biosphere, and earth. To sustain it requires ecological land use, which includes connected wild lands, in which indigenous animals, plant and people can coexist together. Protected core reserves can be set aside for native plants and animals to thrive, without interference. Wildlife corridors can connect reserves such that animals may migrate freely and parks can be established to act as buffer zones between developed and undeveloped areas. Productive rural areas can be re-established through: sustainable agriculture, which does not cause runoff of pollutants into streams and is more in harmony with natural processes; sustainable forestry, which thins rather than clear-cuts stands of trees; sustainable fisheries, which establish quotas, such that species are not depleted; and eco-tourism, to provide natural getaways and education for people that want to learn more about sustainability and to be close to nature. Compact towns and cities, with human scale neighborhoods, green buildings, convenient transit access, ecological infrastructure, and urban growth boundaries will create healthy, vibrant communities.

Economic capital in healthier communities and commerce will tend to expand through synergies with social and environmental capital, thus building prosperity.

ENERGY ENGINEERING FOR DEVELOPING COUNTRIES[14]

Approximately four billion of the six and one half billion people on the planet live in extreme poverty, where a poor sanitation infrastructure results in disease, there is minimal economic productivity, and there is significant economic burden on governments and aid agencies. However, recently, it has been found that given some of the modern necessities such as water wells, electronic communication, and dependable energy, people that are impoverished can quickly and enthusiastically become productive to their communities and not be an economic burden.

The use of small photovoltaic power systems in villages has literally energized villages into minieconomic zones of relative wealth and flow of capital, thus creating self-sufficiency rather than dependence. Just providing electricity for lighting, water pumps, and small power tools can tremendously boost the productive capabilities of a village.

The establishment of cell phone repeaters and wireless infrastructure is much less expensive than running miles of telephone cable. A limited number of cell phones in villages have also promoted prosperity because farmers and merchants can effectively communicate and market their products.

As developing countries continue to develop, water, energy, and communication will be vital to their success. So, energy engineers will be a key part of this process. Many multinational and small companies and financial institutions are tapping into bringing these four billion people into the world of commercial success and out of the world of poverty. So, engineers will play an important part in this next major step toward a sustainable world.

CONCLUSION

Though we are just now seeing the start of a transition to sustainable development, it is clearly economically, ecologically, and socially advantageous to choose this means to success, prosperity, well-being, stability, and peace. We have an abundance of all of the material and energy resources we need. We have the ability, knowledge, and conscience to do the best for humanity, all of life, and the planet. We have the inspiration and insight to transition to a sustainable world. Now all it takes is the willingness to accept change and make the transition to sustainability. The outcome will be a world that is more prosperous and stable than it is now. Sustainable development is a goal to embrace and make part of our daily decisions.

REFERENCES

1. Waage, S. *Ants, Galileo, and Gandhi—Designing the Future of Business Through Nature, Genius, and Compassion*; Greenleaf Publishing: Sheffield S3 8GG, U.K., 2003; [www.greenleaf-publishing.com].
2. Stecky, N. *Introduction to the ASHRAE Greenguide for Leed, Globalcon 2005 Proceedings CD*, https://www.aeecenter.org/store/detail.cfm?id=895&category_id=6.
3. Woolsey, R.J.; Hunter, L.; Amory, L. *Energy Security: It Takes More Than Drilling*, http://www.csmonitor.com/2002/0329/p11s02-coop.html.
4. Woolsey, R.J.; McFarlane, B. Conference Report: *Renewable Energy in America: The Call for Phase II, Session 3: Using Renewable Energy to Meet Our National Security Needs*; December 2004, http://www.acore.org/pdfs/04policy_report.pdf.

5. http://www.insnet.org/ins_headlines.rxml?cust=2&id=1234.
6. http://www.insnet.org/ins_headlines.rxml?cust=2&id=1248.
7. Anderson, C.; Katharine, R. *The Co-Creator's Handbook*, www.globalfamily.net.
8. http://portal.unesco.org/education/en/ev.php-URL_ID=27234&URL_DO=DO_TOPIC&URL_SECTION=201.html.
9. http://www.uspartnership.org.
10. http://www.CorporateRegister.com.
11. Williard, B. *The Sustainability Advantage—Seven Business Case Benefits of a Triple Bottom Line*; New Society Publishers: British Columbia, V0R 1X0, Canada, 2002; [www.newsociety.com].
12. Doppelt, B. *Leading Change Toward Sustainability—A Change-Management Guide for Business, Government and Civil Society*; Greenleaf Publishing: Sheffield S3 8GG, U.K., 2003; [www.greenleaf-publishing.com].
13. http://magma.nationalgeographic.com/ngm/data/2001/07/01/html/ft_20010701.3.html.
14. http://www.csrwire.com/print.cgi?sfArticleId=3709.

Thermal Energy Storage

Brian Silvetti
CALMAC Manufacturing Corporation, Fair Lawn, New Jersey, U.S.A.

Abstract
Often unnoticed, elements of thermal energy storage (TES) technology can be found in many common products. As an engineered system, however, thermal storage has matured into an extensively developed technology primarily for commercial comfort and process cooling. Commercial utility rates and the unbalanced cooling demand of this sector make it an ideal candidate for TES. Although water is the predominant storage material, methods of its use vary across a diverse selection of latent and sensible heat equipment types, each with unique performance and operational characteristics. Additionally, systems can be selected and designed to address any fraction of the cooling load in partial or full storage strategies. In manipulating the contribution of chiller and storage to the cooling load, control logic must be consistent with equipment capabilities and capacities, utility energy and demand rate structures, and integrated load requirements.

INTRODUCTION

Thermal energy storage is one of the world's oldest energy management technologies. For millennia, the mass of shelters has served to moderate extreme temperature excursions by gradually absorbing and releasing heat. And virtually all thermal storage reviews recount the centuries old practice of harvesting ice from rivers and lakes for use throughout the following warmer seasons. Indeed, ice was a major North American export in the 19th century.

One of the unique historical demonstrations of stored cooling occurred on a hot July day in 1620, when the Dutch scientist, Cornelis Van Drebbel, chilled the Great Hall of Westminster Abbey for the benefit of King James I. Van Drebbel reportedly combined snow saved from the previous winter, water, potassium nitrate, and salt in a prescient precursor to the ice cream maker.

There are dozens of more contemporary examples of thermal storage. These include the familiar domestic hot water heater or the household refrigerator icemaker, as well as many lesser known uses, such as in athletic equipment, medical braces and wraps, thermal barriers for fire protection, temperature stabilization of electronic components, and space exploration components.

However, central plant commercial cooling applications have evolved into the most extensively developed sector of thermal storage technology. Professional organizations have published an impressive library of design literature.[1] Equipment and system standards have been developed,[2,3] and manufacturers provide performance data, installation and maintenance instructions, and of course, a wide variety of available products. Other than a broader discussion of storage materials and brief introductions to a selection of additional TES technologies, commercial cooling application and design will dominate the following discussion.

THERMAL STORAGE MATERIALS

Thermal energy is stored by adding or removing heat from a substance. The change experienced by the substance helps define the type of storage system. Although materials such as clathrates and hydrides can store thermal energy through a change in composition, our attention focuses on the more conventional sensible and latent heat properties of materials.

Sensible Heat Storage

When a substance experiences a rise or fall in temperature as heat is added or removed, it is referred to as sensible heat storage. The amount of heat required to raise or lower the temperature of liquid water, its specific heat, is 1 Btu/lb/°F (1 cal/g/°C). That is, 1 Btu will raise or lower the temperature of 1 lb of water 1°F, an unusually high value. High density, another advantageous property of water, improves volumetric efficiency. The density of water is 62.4 lbs/ft^3 (1 g/cm^3). Virtually any material can be used for sensible heat storage, but water, concrete, masonry, rock, earth, and ceramic brick are some of the more common. In fact, the structure itself is sometimes used as a thermal storage medium through precooling or solar heat gain.[4,5]

Keywords: Full storage; Partial storage; Ice storage; Latent heat; Cooling storage; Sensible heat; Thermocline; Chiller; Demand charge; Secondary coolant; Off-peak; TES.

Latent Heat Storage

When liquid water and ice are in thermal equilibrium at 32°F (0°C), the addition or removal of limited quantities of heat does not change the temperature. The addition of heat will change the phase of some of the water from solid to liquid, and the removal of heat will change the phase of some of the water from liquid to solid. This is an example of latent heat of fusion (solid/liquid), the form of latent heat most commonly applied in thermal storage.

Materials employed for thermal storage in this manner are often referred to as phase change materials (PCMs). Water has an extremely high latent heat capacity. It takes the addition or removal of 144 Btu to change the phase of 1 lb of water (80 cal/g) at 32°F (0°C). Water is by far the most widely applied PCM for cooling TES. Its safety, physical properties, high heat capacity, and chemical stability are all important characteristics, but appropriate fusion temperature and negligible cost secure its popularity.

Other latent storage materials have been commercialized or researched over the years for both heating and cooling applications.[6] Sodium sulfate decahydrate (Glauber's salt), one of many inorganic salt hydrates, was perhaps the most intensely studied and typifies both the promise and pitfalls often encountered. Its low cost, high density (specific gravity 1.46), high latent heat capacity (108 Btu/lb, 60 cal/g), and fusion temperature of 89°F (31.7°C) make it apparently ideal for active (mechanically assisted) and some passive heat storage applications. The performance of salt hydrates has proven difficult to perfect because of the tendency to form undesirable hydrates (e.g., sodium sulfate dihydrate), segregate into separate constituents or supercool (undercool). Supercooling is the tendency of most liquids to remain a liquid, even as the temperature drops below the fusion point, until some nucleating event precipitates the formation of solid crystals. Once nucleated, the entire mass of material typically rises rapidly back to its customary fusion temperature.

Mixtures of materials at their eutectic, or lowest freezing point concentration, are also effective storage materials, often employed below 32°F (0°C). For example, 23.3% sodium chloride in water exhibits a uniform freezing temperature at −6°F (−21°C). Some salt hydrates and eutectics are available for thermal storage applications that use a variety of proprietary or patented techniques for improving performance. Rolling drum containers, thixotropic mixtures, gelling agents, mechanical stirring, clever container geometry, microencapsulation, and other chemical additives have all been used to prevent segregation and/or induce nucleation.

Other materials include paraffins, fatty acids, and their eutectic mixtures. These are chemically stable, and the ability to tailor the fusion temperature of paraffins by controlling the carbon chain length is an attractive feature. They typically have lower heat capacities, densities, and thermal conductivities than the water-based materials, and they sometimes experience comparatively high volume changes with phase transition. They are employed in some specialized applications referred to earlier. There have also been several attempts over the years to infuse paraffins into building materials (gypsum wallboard, for instance) to enhance the storage capacity of the building structure.

Although rarely applied, heat transfer to or from many solids at specific temperatures can produce a change in their crystalline structures, another form of latent heat. Some materials will even undergo a series of structural changes (e.g., hexatriacontane) at different temperature plateaus.

TES FOR COOLING APPLICATIONS

Benefits

Many factors encourage the use of thermal storage. Most applications, not only the classic examples like churches, dairies, and theaters that exhibit brief but also intense cooling loads, benefit from substantial reductions in refrigeration equipment and electrical infrastructure. But the principal basis for many TES benefits is energy cost reduction, particularly with the current emphasis on Leadership in Energy and Environmental Design (LEED®) certification.

In addition to the recognized benefits of avoiding the construction of additional generation and transmission facilities, there is also substantial evidence that power generated off-peak is produced more efficiently and with fewer emissions than the peaking power it replaces.[7]

Engineers often install excess capacity to meet unexpected cooling loads or to compensate for mechanical failure, an inefficient and expensive practice. Rather than invest in chiller capacity that will rarely be used, and may produce even higher electrical demand, the same objectives can be achieved with TES while realizing continuing economic benefits. A selection example is included under "System Design."

Commercial Utility Rates and TES

The commercial customer utility bill is usually comprised of two main components. The first, the energy charge, is based on the amount of electrical energy consumed, or kWh. This energy charge is often reduced during off-peak hours, providing the first incentive for TES. The second component is referred to as a "demand" charge and is based on the rate (kW) at which a customer consumes energy, often for the highest 15-min demand period during on-peak hours. A demand ratchet may extend the impact of high demand to subsequent months. The demand component may be incorporated in an energy charge that is reduced for customers with a flatter electrical demand profile, rather than explicitly identified.

To illustrate, a customer that operates a 100-W light bulb for 10 h will consume the same amount of electric energy as a customer who burns ten 100-W light bulbs for 1 h. However, the customer who consumes that energy over the 10-h period will pay one-tenth of the demand-related utility cost. It is quite common for a commercial customer's demand charges to contribute more than 50% of the utility bill.

Refrigeration equipment can constitute 40% or more of a commercial building's peak electric demand, with little or no nighttime cooling load, representing a significant opportunity for energy cost savings with TES. In a very real sense, TES applications that reduce on-peak electrical usage are storing electricity as much as heating or cooling capacity.

Cooling Storage Equipment

Cooling storage devices encompass a wide variety of equipment designs with unique characteristics. Although far from comprehensive, cool storage systems can usually be classified within one of the following groups, as adopted by most manufacturers and ASHRAE.[8] Because storage systems must meet not only the peak instantaneous load (tons), but also provide the integrated cooling capacity over the design period, an essential measure of capacity is the ton-hour (0.0127 GJ).

Ice Storage-HX Coil Builds and Melts Ice

Fig. 1 illustrates a typical design. The complete assembly is usually factory manufactured and provided in thermally insulated, modular units that can be combined to provide the needed capacity (Fig. 2). Cylindrical plastic tanks and rectangular galvanized or stainless steel tanks are available. Plastic tanks provide water containment, while metal tanks include an additional flexible or metal liner to achieve water containment.

An antifreeze solution, usually 25% ethylene glycol/75% water[9] and referred to as a secondary coolant, is circulated through the tubular heat exchanger (HX) that is distributed throughout most of the tank's internal volume. The HX is completely immersed in water that usually occupies about 80% of the tank volume. The remaining space is reserved for the HX and to allow for the expansion of the water as it freezes, decreasing in density by approximately 9%. The expansion of the water into the space above the HX is often used as a measure of ice inventory. Heat exchangers are constructed of plastic, usually polyethylene or polypropylene, or galvanized steel. Copper heat exchangers circulating refrigerants are a more recent development in TES systems with smaller capacities. The spacing between the tubes of the HX is determined by the diameter of the tubes themselves, desired rates of charging (freezing) and discharging (melting), and heat transfer properties of the HX and coolant.

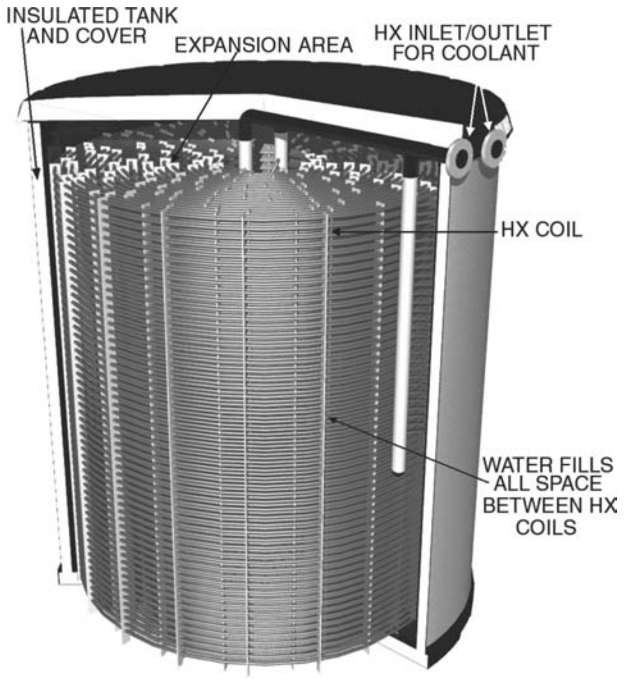


Fig. 1 Modular ice storage tank—charge and discharge through HX. (Courtesy CALMAC Manufacturing and Nihon Spindle.)

Note that the water in the tank is not hydronically connected to the building-cooling loop and does not circulate. All of the heat transfer is accomplished by circulation of the coolant through the HX. The HX usually operates as part of the pressurized, closed building loop while the tank itself is at atmospheric pressure.

Standard chillers (excluding absorbers) provide the cooling source for freezing the water. During the charging cycle, the chiller cools the secondary coolant to a range of perhaps 20°F (−6.7°C) to 27°F (−2.8°C), the lower temperatures reached at the end of the charge cycle. Coolant is circulated through the HX until all the water between the HX tubes is frozen (Fig. 3).

During the discharge, the same coolant is circulated between the cooling load and the storage tank. As the ice surrounding the HX is melted, the 32°F (0°C) liquid water/ice surface gradually retreats from the HX tube surface (Fig. 3). Therefore, the performance, as reflected in the temperature of the coolant leaving the HX, varies as the ice is melted. Some manufacturers publish charts to describe the variable performance due to this characteristic. This design sometimes employs alternate PCMs.

Ice Storage-HX Coil Builds Ice-Surrounding Water Melts Ice

This design shares some features of the internal melt system. Tanks are typically large, site-built, rectangular concrete structures that operate at atmospheric pressure (Fig. 4). The HXs are usually galvanized steel coils,

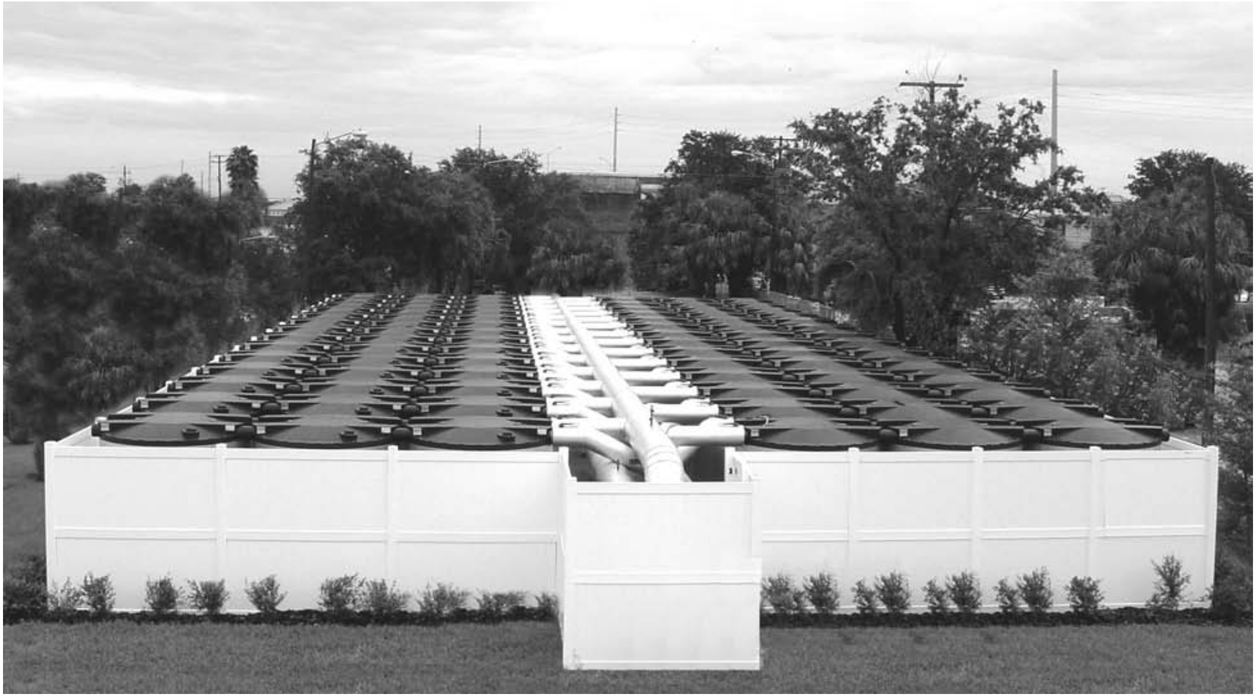


Fig. 2 14,000 tn-h, 5000 ft^2 modular ice storage system, cooling 1,000,000 ft^2 district system. (Courtesy CALMAC Manufacturing.)

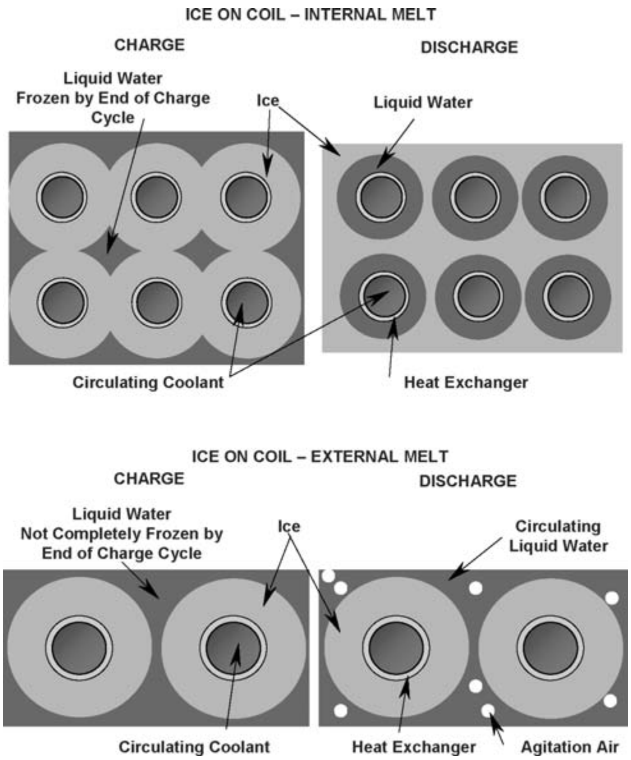


Fig. 3 Comparison of charging and discharging processes from interior and exterior of ice on coil.

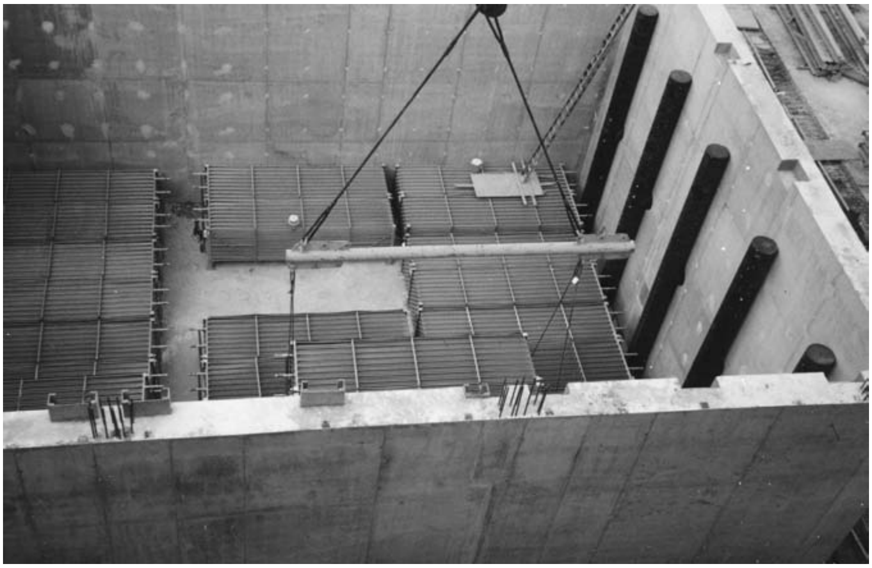


Fig. 4 50,000 tn-h district cooling storage system under construction. (Courtesy CALMAC Manufacturing.)

similar to the internal melt coil, but often with wider tube spacing.

Circulating cold coolant through the HX coils, as in the internal melt design, forms the ice. The major difference is that some liquid water is always left surrounding the ice formed on the HX tubes (Fig. 3). This water is the heat transfer fluid for the discharge cycle, and ice thickness must be carefully controlled to ensure that a flow passage is always available for the circulating water while still achieving the required ice inventory. Because the liquid water is in continuous direct contact with the ice, it is available at consistently low temperatures and high rates of discharge.

Various methods are used to interface the atmospheric pressure tank with a pressurized building-cooling loop when necessary. For instance, an intermediate HX or pressure sustaining valves (see Fig. 5 for stratified chilled water) can be placed between the two systems.

The velocity of the water within the tank is extremely low. Compressed air is typically distributed through separate piping located under the bottom layer of coils to augment convection and promote even melting and freezing of the ice throughout the often considerable dimensions of the tank.

The refrigeration equipment often contributes to the cooling load during the discharge cycle (see "System Design"), but separation of the chiller coolant (antifreeze solution) from the water loop must be maintained. Rather than simply continue to circulate the coolant through the ice-encased storage coils, improved efficiency can be realized by adding an additional HX between the chiller coolant loop and the building water loop. Sometimes the actual refrigerant is redirected to a separate evaporator where it cools the circulating water.

Although the chiller/coolant method of ice-build has become common, direct expansion or liquid-overfeed of refrigerants, circulated directly through the ice coils, is applied in certain process applications. Water is the only PCM currently used in external melt systems.

Ice Storage-PCM in Sealed Containers or Encapsulated

These systems also employ standard chillers circulating a secondary coolant. However, the PCM, usually water, is contained within sealed plastic containers. Insulated tanks are filled with the containers, and coolant circulates through the tank and around the containers.

Tank configurations include horizontal or vertical cylinders, both pressurized and at atmospheric pressure. The useable volume of PCM depends on the shape of the container, the geometry of the tank, and whether the containers are individually positioned or allowed to pack randomly.

Many container shapes have appeared over the years, including cylindrical tubes and flat trays, but current products are mainly spherical. Additional features, intended to enhance heat transfer or accommodate expansion of the water, include flexible surface geometry, internal voids, or collapsible internal chambers.

Encapsulated systems are available with alternate PCMs, usually salt hydrates and eutectics, but other materials are possible. Supercooling must be prevented by including a nucleating agent or device.

Ice Storage-Harvester

Refrigerant is directed through a collection of vertical plates or tubes while water constantly flows around the external surface. Ice accumulates on the surface, usually to a thickness less than ½ in. Repositioning an arrangement of valves introduces compressed, hot refrigerant gas into

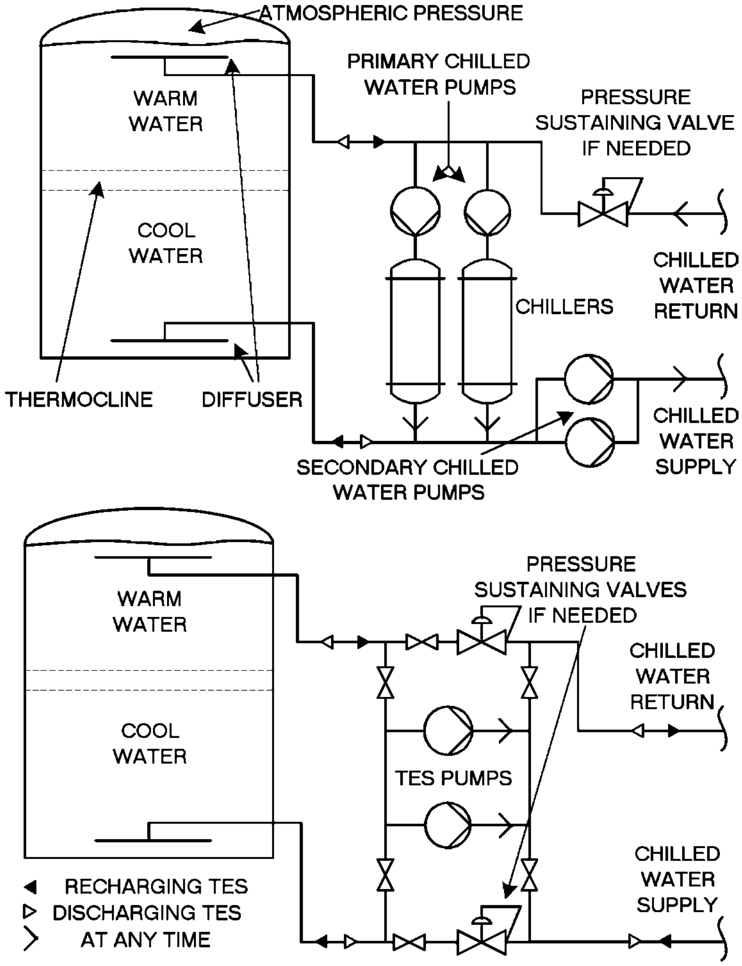


Fig. 5 Chilled water storage piping schematics Top: TES near chiller water plant. Bottom: TES near or remote from chilled water plant. (Designs courtesy The Coolsolutions Company.)

the plate. The ice separates from the surface and drops into an atmospheric pressure steel or concrete tank below. This procedure is repeated sequentially over a series of many plates or tubes, so that the process is essentially continuous.

An alternate approach circulates coolant or refrigerant around an HX that has water combined with a low concentration of glycol flowing on the opposite surface. A combination of mechanical disturbance and the fluid's reduced tendency to adhere to a surface when frozen eliminate the need for hot gas defrost.

Harvesters are less common in comfort conditioning applications today but are still used in process and industrial applications. As in the external melt design, the direct contact between the circulating water and the ice provides low temperatures and high discharge rates.

Stratified Chilled Water Storage

Chilled water is a sensible heat storage method. The available capacity is proportional to the volume of water and the temperature change it experiences. Cost considerations dictate that the same tank, usually a steel or concrete vertical cylinder, be used to store both the warm water returning from the cooling load and the cold water circulated to the cooling load. It is critical that the warm and cool water are prevented from mixing within the tank.

The preferred approach that has evolved is the simple stratified water system. Because warm water is less dense than cooler water, it will float on cooler water below. However, the density of water at 58°F (14.4°C) is only about 0.05 lb/ft^3 (0.8 kg/m^3) less than water at 40°F (4.4°C). Maintaining separation is now commonly achieved through careful design of the water diffusers located in the top and bottom of the tank. The goal is to prevent disruption of the stable but delicate density relationship as water enters and leaves.

Two nondimensional flow parameters have usually been accepted as crucial to proper performance. The Froude number is the ratio of inertial to gravitational forces (related to the Richardson number), and the

Reynold's number is the ratio of inertial to viscous forces. Research, with sometimes inconsistent results, is still evaluating the influences of these and other parameters.[10] Designers, using a combination of analytical and empirical approaches, have nonetheless arrived at a successful design process.

Fig. 5 represents chilled water storage systems part way through a charge or discharge cycle. During charge, flow exits the top of the tank, is cooled by the chiller, and returns to the tank bottom. During discharge, the flow reverses, and water exits the bottom of the tank, is distributed to the load, and returns to the top. Note the area between the warm and cold layers. Known as a thermocline, this short cylinder of water, usually 1–3 ft thick, exhibits a vertical temperature gradient between the cold and warm water temperatures.

Because the maximum density point occurs at 39.2°F (4°C), water is usually cooled to no less than 40°F (4.4°C), and chiller efficiencies are very good during the charging cycle. Additives are sometimes employed to lower the temperature of maximum density and increase the storage capacity. Increasing the system return temperature also expands the storage capacity. However, the water must return to the tank from the cooling system at the design temperature because any reduction represents lost capacity.

Large chilled water storage systems are very economical. Examples include a district cooling application with a 17.6 million gallon (66.6 million liter) tank, 160,000 tn-h total capacity, and a peak discharge rate of 20,000 tn.

Cool Storage System Design

All equipment types require some unique design elements, but an ice storage system will introduce most of the important issues.

Load Profile

The first step is to determine the cooling load profile for the entire design day. In a nonstorage system, the peak load is the primary influence on chiller selection. For storage systems, both the peak load and the total combined load for the design period will determine chiller and storage capacities. Furthermore, chillers may operate at different capacities at different times of the day, and storage performance may vary with system geometry, control sequence, storage inventory, and system temperatures. For this reason, ARI has published a format for specifying operating temperatures, modes, loads, etc. on an hour-by-hour basis.[11]

An assumed load profile for a typical office building is depicted in Fig. 6 (top). The peak load is 1000 tn and the total integrated cooling load over a 12-h period is 10,000 tn-h.

Chiller and Storage Capacity

Because all of the cooling originates with the chiller, we first calculate chiller capacity by equating the full load chiller operating hours to the total cooling load. The calculations are quite simple, requiring only careful assignment of the hourly operating modes, consistent with the utility tariff, building occupancy, and the planned strategy.[12] There are a number of options available to the designer. Possible modes of operation for any particular time include charging (ice-making), charging with cooling, cooling with a combination of storage and chiller, cooling with chiller only, cooling with storage only, and system off.

$$\text{Total ton hours} = \text{Chiller day capacity} + \text{Chiller ice capacity} \quad (1)$$

Full storage is the easiest to calculate and the simplest to implement. It also provides the greatest operating cost savings, but the highest initial investment. The entire on-peak cooling load is provided from storage, and the chiller only operates to produce ice at night, off-peak.

In this simplest of cases, the day capacity is zero, because the chiller is off. The ice-making capacity, at this point unknown, will be less than the chiller's nominal daytime capacity. However, the relative daytime and ice-making capacities are easily determined for a particular chiller category. A centrifugal chiller, for instance, might produce 70% of its nominal capacity when in the ice-making mode, and 100% during the day. We usually want to describe the required chiller capacity in terms of its conventional rating and will therefore include a factor of 0.7 for hours operated at ice-making conditions. This is a capacity modifier, not an efficiency adjustment. Whether the efficiency is the same, better, or somewhat worse during charging will depend on a number of factors, such as nighttime condenser temperatures, arrangement and type of components, and choice of air or water cooled chiller.[13] With 12 h available for charging storage, Eq. 1 becomes:

$$10{,}000 \text{ tn h} = 0 + (\text{NC} \times 12 \text{ h} \times 0.7)$$

$$\text{NC} = \frac{10{,}000 \text{ tn h}}{12 \text{ h} \times 0.7} = 1190 \text{ tn}$$

where NC is the nominal chiller size. Of course, the required storage capacity is the entire 10,000 tn-h.

Another more common approach is to operate the chiller throughout the day, producing ice at night and directly contributing to the load during the day. At its limit, this results in the smallest possible chiller capacity, and equipment costs are often competitive with nonstorage systems.

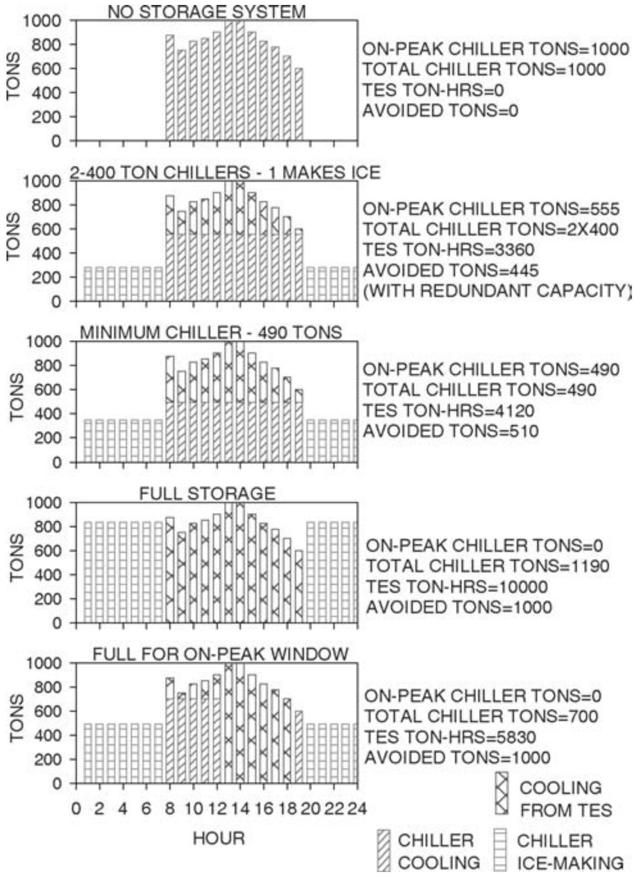


Fig. 6 Alternative TES selections for the same cooling load profile.

$$10{,}000 \text{ tn h} = (\text{NC} \times 12 \text{ h} \times 1) + (\text{NC} \times 12 \text{ h} \times 0.7)$$

$$\text{NC} = \frac{10{,}000 \text{ tn h}}{(12 \text{ h} + 8.4 \text{ h})} = 490 \text{ tn}$$

The amount of storage becomes:

$$\text{Storage ton hours} = \text{NC} \times \text{Ice making hours} \times \text{Capacity factor} \quad (2)$$

$$\text{Storage ton hours} = 490 \text{ tn} \times 12 \text{ h} \times 0.7 = 4118 \text{ tn h}$$

Note that this partial storage chiller is only 49% of the peak load and 41% of the full storage selection. Storage capacity is also reduced to 41% of the full storage requirement.

There is any number of intermediate selections. For instance, a common conventional design might include three 400-tn chillers for our 1000-tn building. An alternative is to select two 400-tn chillers and storage equal to the ice-making capacity of one of the machines, or about 3360 tn-h for our example. Similar redundancy is provided in the event of a chiller failure, with the added benefit of permanent operating cost savings.

Other multiple chiller options are possible. Certain utilities experience high, but relatively short, afternoon demand peaks, and a common alternative in these areas treats a 4–6 h period of the day as full storage. Other applications, churches for instance, may operate on multiple-day cycles, or industrial applications may experience a series of brief but intense cooling loads over the period of several hours. Fig. 6 depicts some possible selections and the on-peak chiller reductions that are possible with the more common daily cycles.

System Layout and Control

The final step is to construct a simple schematic representation of the proposed system and verify the control logic. A common partial storage design places the storage and chiller in series (Fig. 7). Our example positions the chiller upstream of the storage relative to the cooling load, but the reverse is sometimes preferred. Remember that in the internal melt system, the only fluid circulating is the freeze-protected coolant; the storage water/ice never leaves the storage tanks.

Note the modulating three-way valve located at the storage system. This valve automatically controls coolant flow through the storage tanks to maintain the desired

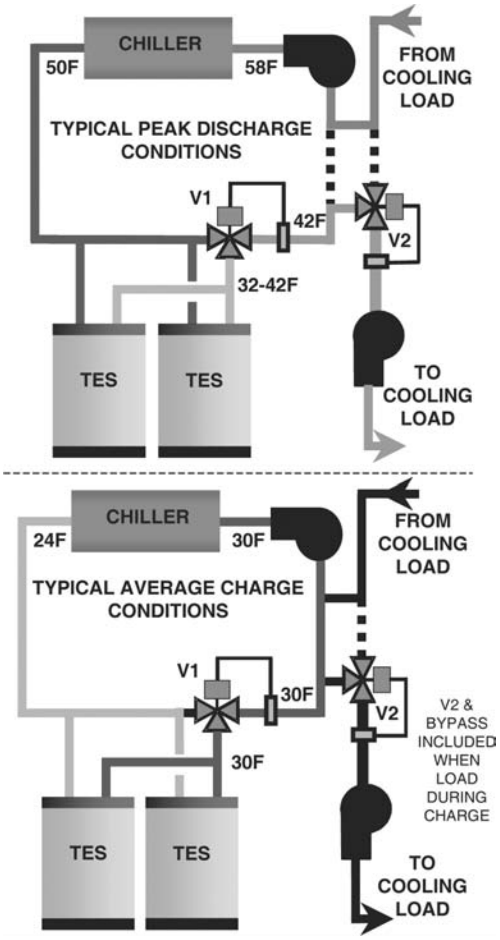


Fig. 7 Ice storage system schematic—charge and discharge.

coolant temperature throughout the discharge period as ice is melted.

During charging, the chiller operates at the lower ice-making temperatures. The three-way valve directs all flow through storage. With no direct cooling needed during the ice-making period, the load can simply be bypassed. If there is a cooling load during ice-making, a second three-way valve and bypass is sometimes included (dashed lines) so that warm coolant returning from the load can be blended with the cold ice-making coolant before being sent to the cooling coils. The ability to efficiently meet small nighttime loads is often seen as a major storage advantage. A separate chiller often serves substantial nighttime load.

Discharge logic is dependent on the utility rate, equipment limitations, cooling loads, and the level of on-site supervision. The chiller can simply be set to provide the design supply temperature (e.g., 42°F [5.6°C]). Storage will only be discharged if the smaller chiller cannot meet the load. The amount of ice melted each day is minimized, referred to as chiller priority. Alternately, on cooler than design days, the chiller leaving temperature can be raised to increase the load on storage (storage priority) and take advantage of additional demand or energy cost savings. Of course, the danger is that an error in the estimated total cooling load may exhaust storage, leaving only an undersized chiller to carry the load later in the day. Even unsophisticated systems, however, successfully employ this strategy by incorporating adequate margins in the level of chiller loading.[14]

Verify Initial Assumptions

Once the control and operating sequences have been established, return to the original selection calculations to insure that chiller and/or storage contribution, device operating temperatures, and modes of operation remain consistent with original assumptions. There are many alternative designs. Some include parallel flow arrangements, additional heat exchangers, or chillers. By applying the proposed logic on an hour-by-hour basis, the designer is assured of reliable performance.

BRIEF REVIEW OF OTHER TES OPTIONS

There are, of course, many other interesting and promising areas of TES technology. In some cases, precooling of the building structure and contents provides significant storage.[5] Research in this area is identifying and quantifying relevant design parameters, such as the location and type of auxiliary mass, effects of space dimensions, control methods, and potential benefits.

Underground thermal storage systems are used for both cooling and heating. Ground water from aquifers (ATES) is transferred between a warm well and cold well, extracting or adding heat either directly or through a heat pump. Alternatively, a closed loop array of vertical heat exchangers is placed into boreholes (DTES). Some issues that influence the choice or design of these systems include the balance between seasonal heating and cooling loads of the building, local geology, topography, hydrology, extraction and reinjection flow rates, and impact on neighboring water use (wells).[15]

Some winter-peaking utilities promote off-peak heat storage. A common design distributes an array of electric resistance heaters throughout a matrix of high-temperature ceramic (or other refractory) bricks within an insulated cabinet. Temperatures in excess of 1400°F (760°C) can be reached, providing large storage capacities in a small volume utilizing only sensible heat effects. A fan circulating air through passages within the storage device may assist discharge.[16]

Heat storage, combined with passive solar techniques, can be very effective if properly applied with regard to local climate, space geometry, etc. The Trombe wall, popularized in the mid-1960s, provides insight into several solar heating principles,[4] and, of course, the simple hot water tank is a common storage vessel for active solar systems.

CONCLUSION

Progress continues in many areas of TES. Unique applications, like combustion turbine inlet air-cooling that recovers capacity normally lost during hot weather, are continually being developed.[17] Efforts are currently underway to describe more precisely TES performance in energy modeling software, motivated largely by increasing interest in the U.S. Green Building Council (USGBC) LEED®[18] certification program. Thermal storage is a uniquely diverse area of energy technology with both a long history of reliable use and the emerging potential of new and interesting developments.

REFERENCES

1. Dorgan, C.E.; Elleson, J.S. *Design Guide for Cool Thermal Storage*; ASHRAE: Atlanta, GA, 1993.
2. ANSI/ASHRAE/Standard 150-2000. *Method of Testing the Performance of Cool Storage Systems*, ASHRAE: Atlanta, GA, 2000.
3. ARI Standard 900-98. *Thermal Storage Equipment Used for Cooling*, Air-Conditioning and Refrigeration Institute: Arlington, VA, 1998.
4. Torcellini, P.; Pless, S. *Trombe Walls in Low-Energy Buildings: Practical Experiences (Preprint)*; National Renewal Energy Laboratory: Golden, CO, 2004.
5. Braun, J.E. Load Control Using Building Thermal Mass. J. Solar Energy Eng. **2003**, *125*, 292–301.
6. Zalba, B.; Marin, J.; Cabeza, L.; Mehling, H. Review on thermal energy storage with phase change: materials, heat transfer analysis. Appl. Thermal Eng. **2003**, *23*, 251–283.
7. Tabors Caramanis & Associates. *Source Energy and Environmental Impacts of Thermal Energy Storage*, California Energy Commission: Davis, CA, 1995.
8. Elleson, J.; Kooy, R.; Bahnfleth, W. *2003 ASHRAE Handbook, HVAC Applications, Chapter 34 Thermal Storage*, ASHRAE: Atlanta, GA, 2003.
9. Gatley, D.P. *Cool Storage Ethylene Glycol Design Guide*, EPRI TR-100945; Electric Power Research Institute: Palo Alto, CA, 1992.
10. Bahnfleth, W.P.; Musser, A. Thermal performance of a full-scale stratified chilled-water thermal storage tank. ASHRAE Trans. **1998**, 377–388.
11. ARI Guideline T-2002. Specifying the Thermal Performance of Cool Storage Equipment, Air-Conditioning and Refrigeration Institute: Arlington, VA, 2002.
12. Silvetti, B. Application fundamentals of ice-based thermal storage. ASHRAE J. **2002**, *February*, 30–35.
13. MacCracken, M.M. Thermal energy storage myths. ASHRAE J. **2003**, *September*, 36–43.
14. Elleson, J.S. *Successful Cool Storage Projects: From Planning To Operation*; ASHRAE: Atlanta, GA, 1997.
15. http://www.worldenergy.org/wec-geis/publications/default/tech_papers/17th_congress/4_1_16.asp (accessed June 2005).
16. http://www.electricenergyonline.com/IndustryNews.asp?m=&id=38585 (accessed July 2005).
17. Andrepont, J.S. Thermal energy storage technologies for turbine inlet cooling. Energy-Tech. **2004**, *August*, 18–19.
18. *LEED Green Building Rating System For New Construction & Major Renovations (LEED-NC) Version 2.1*, U.S. Green Building Council: Washington, DC, 2002.

Thermodynamics

Ronald L. Klaus
VAST Power Systems, Elkhart, Indiana, U.S.A.

Abstract
The foundational ideas of thermodynamics, especially the implications of its two main laws, are presented, and their implications for energy engineering are discussed. Special attention is given to the implications of the second law in the design of processes that are thermodynamically efficient. The application of thermodynamics to practical engineering problems requires accurate estimation of thermodynamic properties of fluids. A survey of the methods used for such estimations is also presented.

NOMENCLATURE

All thermodynamic properties have the units of energy or energy per unit time except where otherwise specified.

C_p	Heat capacity at constant pressure (energy/temperature)
e^{Ch}	Chemical exergy
e^{Ph}	Physical exergy
G	Gibbs free energy, or Gibbs function
H	Enthalpy
M	General thermodynamic property
M_i	Partial molal property of specie "i"
n	total number of moles
n_i	number of moles of specie "i"
P	Pressure (force/area)
P^V	Vapor pressure (force/area)
Q	Heat or heat flux added to a process (energy or energy per unit time)
Q_{ideal}	Heat or heat flux added to a completely reversible process (energy or energy per unit time)
R	Universal gas constant (energy/temperature)
S	Entropy (energy/temperature)
T	Temperature
T_0	Reference temperature or temperature of the surroundings
U	Internal energy
V	Volume (length^3)
W	Work or power produced by a process. Usually does not include work done by or on fluids entering or leaving the process (energy or energy per unit time)
W_{ideal}	Work or power produced by a perfectly reversible process (energy or energy per unit time)
W_{lost}	Lost work or power (energy or energy per unit time)
W_{total}	Total work or power produced by the process (energy or energy per unit time)
x_i	mole fraction of specie "i"
x	vector containing all mole fractions
yH_2O	Mole fraction of water vapor in a vapor phase
Z	Compressibility factor
Δ	Difference in the value between two states or differences between the properties of outlet and input streams
ΔH_f	Enthalpy of formation
ΔS_f	Entropy of formation (energy/temperature)

Subscripts

f	Formation
o	Base or reference state
C	Cold sink
H	Hot source
rev	Reversible

Superscripts

*	Ideal gas or ideal gas state
id	Ideal solution

Keywords: Combustion; Energy; Enthalpy; Entropy; Exergy; Gibbs- free energy; Ideal gas; Irreversibility; Lost work; Power generation; Phase equilibrium; Thermodynamic properties.

INTRODUCTION

Thermodynamics is concerned with the interaction between matter and energy on a macroscopic level. It is therefore the basic science that underlies the engineering of the use of energy in all its forms. It is governed by two basic laws with far-reaching implications. The first concerns the conservation of energy, and the second puts limits on the amount of energy that can be converted into work over and above what the first law might be thought to imply.

The actual performance of practical process equipment can be predicted by application of these two laws. That, in turn, depends on the ability to estimate the thermodynamic properties of the streams that enter and leave this equipment. The exact equations needed to calculate these properties can be derived from the formalism of thermodynamics. Certain simplifications, such as the ideal

Ther–Und

gas law or the assumption of an equation of state, are usually required to make numerical estimates of these properties.

THE FIRST LAW OF THERMODYNAMICS

Formulation

The first law of thermodynamics is a statement of the principle of conservation of energy for a closed system—namely, one in which there is no transfer of mass across the system boundaries.

$$\Delta U = Q - W_{\text{total}} \quad (1)$$

This equation ignores certain effects such as differences in fluid velocities, which become important only at very high velocities, and differences in fluid elevations, which become important only when there are greatly varying elevations. U is a thermodynamic property called internal energy. It accounts for energy due to the motion of molecules in a fluid and the exchanges of energy between them. The Δs refer to differences in the property value between the final and initial states of the system. U, Q, and W all have units of energy. However, this equation can also have another interpretation for open flow systems at steady state. Such systems have input and outlet streams, and possibly an exchange of heat or work with the surroundings (Fig. 1). However, the stream flows and property values throughout the system are all considered constant over time. In this case, the first law is a rate equation. ΔU is the difference between the sum of U times the mass flow rate for all the output streams minus that same sum for all the inlet streams. Q and W are rates of energy transfer. The arrows in Fig. 1 indicate the direction of the flow of work, W, and heat, Q, when these quantities are positive.

When applying Eq. 1 to a flow process, W_{total} includes the work done in moving fluids into and out of the process. To eliminate these effects, another thermodynamic property, namely enthalpy, H, is defined,

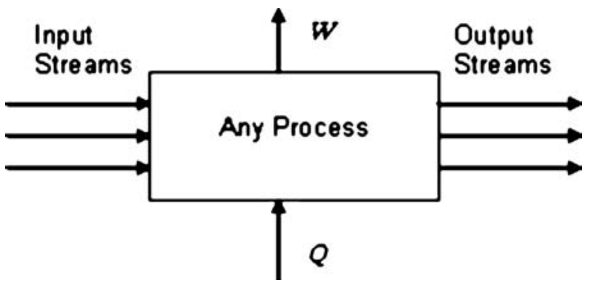


Fig. 1 Schematic representation of any thermodynamic process.

$$H = U + PV \quad (2)$$

where P is pressure and V is volume.

For a flow process at steady state, this changes the first law to

$$\Delta H = Q - W \quad (3)$$

In Eq. 3, W is the work produced by the process apart from that involved in moving fluids into and out of the process. It is often called the shaft work.

In these equations, it is important to make a distinction between intrinsic and extrinsic properties. The former are properties per unit mass or per mole, whereas the latter are total values of the property for the mass flow of the given process stream. These and other thermodynamic relationships can be applied to either, provided that they are applied consistently.

U, H, P, and V are examples of thermodynamic state functions—that is, their intrinsic values depend only on the state of the fluid (usually its temperature, pressure, and composition) and not on the process used to bring the fluid to that state. On the other hand, quantities like Q and W are not state functions. When a fluid makes a transition from one state to another, the values of Q and W associated with that transition are dependent on the "path" over which that transition was made.

Enthalpies of Formation and Reaction

Enthalpy cannot be calculated in an absolute sense. It requires the definition of a reference state, at which it is taken to be zero. Enthalpies in all other states are referenced to this state. The most rigorous definition of a reference state is to take it as comprising the constituent atomic species in their normal molecular configurations at some temperature and pressure—typically 25°C and 1 atm. Normal molecular configuration means the normal configuration the atomic molecules take in nature, such as O_2, H_2, N_2, etc. Thus, these molecules would be taken to have zero enthalpy at 25°C and 1 atm. The phases of these elemental species also have to be stated. They are generally taken to be the normal phase of the element at the stated conditions. The reference states for gases are usually taken to be in their "ideal gas state" (ig).[1]

With these ideas as a starting point, the enthalpies of compounds can be built up by considering real or imaginary reactions through which they are formed from their constituent elements. For example, the enthalpy of formation of methane would be the enthalpy change associated with the following reaction:

$$C(s) + 2H_2(ig) = CH_4(ig) \quad (25°C, 1 \text{ atm})$$

where s indicates solid and ig indicates the ideal gas state. Such chemical reactions are usually accompanied by the absorption or release of a certain amount of heat. If the reaction is carried out at constant temperature and pressure,

Table 1 Enthalpies of formation of some common compounds

Chemical specie	Formula	State	MW	ΔH_f (T=298.15°C)	ΔH_f (non dim.)
Methane	CH_4	IG	16.04	−74.87	−32.968
Ethane	C_2H_6	IG	30.08	−83.8	−36.901
Propane	C_3H_8	IG	44.10	−104.7	−46.104
n-Octane	C_8H_{18}	IG	114.23	−208.4	−91.767
n-Octane	C_8H_{18}	L	114.23	−250.3	−110.217
Carbon dioxide	CO_2	IG	44.01	−393.51	−173.278
Carbon monoxide	CO	IG	28.01	−110.53	−48.671
Water	H_2O	IG	18.02	−241.826	−106.486
Water	H_2O	L	18.02	−285.830	−125.863

Units of ΔH_f=kJ/gmole; Non-dimensionalization, $\Delta H_f/RT_0$ where T_0=273.15 °K.

and if no work is done, according to the first law, this heat is the enthalpy difference of the reaction and defines the standard enthalpy (or heat) of formation of the resulting compound.

Standard enthalpies of formation are widely tabulated. An excellent source is the collection by the National Institute of Standards and Technology (NIST).[2] Enthalpies of formation for some common species are given in Table 1. They are given both in the units that appear in the NIST collection and in nondimensionalized form. The latter permits their use in any system of units simply by multiplying by RT_0 in the desired units. T_0 is taken to be 0°C (=273.15 °K=459.67 °*R*), and *R* is the Universal Gas Constant. Note that °*R* represents degrees Rankine and °K represents degrees Kelvin.

Many different units are used in energy calculations. The values of some useful conversion factors and the values of the universal gas constant appear in Table 2.

These enthalpies of formation can be combined to produce enthalpy changes for any chemical reaction whose species have known enthalpies of formation. For example, in the combustion reaction

$$CH_4 + 2O_2 = CO_2 + 2H_2O$$

the enthalpy change may be calculated as follows:

$$\Delta H = \Delta H_f(CO_2) + 2\Delta H_f(H_2O) - \Delta H_f(CH_4) - 2\Delta H_f(O_2)$$
$$= (-393.51) + 2(-285.830) - (-74.87) - 0$$
$$= -890.3 \text{ kJ/gmole}$$

If ΔH is positive, the reaction is called endothermic, whereas if it is negative, it is called exothermic. Because this particular reaction is a combustion reaction, the absolute value of its enthalpy change is also called the enthalpy change (or heat) of combustion.

Application to Combustion

The goal of a power cycle is to convert the chemical energy of a fuel into usable mechanical energy that can produce useful shaft work or drive a generator to produce electricity (Fig. 2). What one hopes to do is to convert as much of the fuel's chemical energy into useful work. The hope for this ideal is reflected in what is called the heating value of the fuel. It is taken to be the negative of the enthalpy change of the combustion reaction through which a fuel is completely oxidized at the so-called ISO (International Organization for Standardization) conditions—namely, 15°C and 1 atm pressure.[3] Thus, it is the negative of the enthalpy change of the following kind of reaction:

$$\text{Fuel} + aO_2(\text{ig}) \rightarrow bCO_2(\text{ig}) + cH_2O(\text{ig or l}) + \text{Any other products}(15°\text{C}, 1\text{ atm})$$

where *a*, *b*, and *c* are coefficients that depend on the particular fuel.

It is unfortunate that ISO conditions differ from the reference state at which enthalpies of formation are usually tabulated. Nevertheless, the latter can be adjusted to ISO conditions, and values for some common fuels appear in Table 3.

Heating Values, Heat Rates, and Power Cycle Efficiency

One area of confusion about heating values comes because of the choice of the phase of the water produced as a combustion product. If the product water is considered to be a liquid, its enthalpy will be lower than that if it is considered to be a vapor. The former choice leads to

Ther–Und

Table 2 Some useful conversion factors and constants

Type	Value	Conversion
Length	1 m	3.28084 ft
		39.3701 in.
Mass	1 kg	2.20462 lbm
Force	1 N	1 kg m/s^2
		0.224809 lbf
Pressure	1 bar	10^5 N/m^2
		10^5 Pa
		0.986923 atm
		14.5038 psia
Pressure	1 atm	14.6960 psia
Volume	1 m^3	35.3147 ft^3
Density	1 g/cm^3	62.4278 lbm/ft^3
Energy	1 J	1 N m
		1 m^3 Pa
		10^{-5} m^3 bar
		10^{-3} kW s
		9.86923 cm^3 atm
		0.239006 cal
		5.12197×10^{-3} ft^3 psia
		0.737562 ft lbf
		9.47831×10^{-4} Btu
Power	1 kW	10^3 J/s
		239.006 cal/s
		737.562 ft lbf/s
		0.94783 Btu/s
		1.34102 hp

Values of the universal gas constant, R

8.314 J/gmole K = 8.314 m^3 Pa/(gmol K) = 83.14 cm^3 bar/(gmole K)

82.06 cm^3 atm/(gmole K)

1.987 cal/(gmole K) = 1.986 Btu/(lbmole R)

0.7302 ft^3 atm/(lbmole R) = 10.73 ft^3 psia/(lbmole R)

1545 ft lbf/(lbmole R)

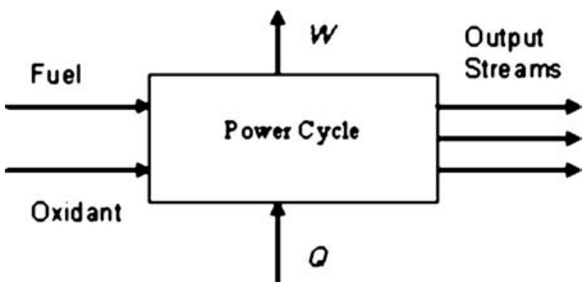


Fig. 2 Schematic representation of a power cycle.

a higher heating value (HHV), whereas the latter leads to a lower heating value (LHV). If the chemical composition of the fuel is known, the heating value is the mole fraction average of the heating values of the individual constituents divided by the average molecular weight of the fuel. Where the composition of the fuel is not known, accurate heating values are determined experimentally.

The heat rate is one often-tabulated measure that is used to describe the effectiveness of a machine—such as an internal combustion engine or a gas turbine—used to convert the chemical energy of a fuel into usable mechanical or electrical energy. Several definitions go into it.

$$\begin{aligned}\text{Heat Energy Input(Btu/s)} &= \text{First Law Available Chemical Energy} \\ &= \text{Heating Value of a Fuel (Btu/lb)} \\ &\quad \times \text{Fuel flow rate (lb/s)}\end{aligned}$$

$$\begin{aligned}\text{Heat Rate(Btu/kWh)} &= \text{Heat Energy Input (Btu/s)} \\ &\quad \times 3600\ \text{(s/h)} / \text{Power Produced (kW)}\end{aligned}$$

The heat rate is based on a particular choice of either HHV or LHV of the fuel. The higher the heat rate, the more fuel is being used per kW of power produced and therefore the less efficient the machine. The heat rate is directly related to the efficiency.

$$\text{Efficiency} = \frac{\text{Power Output (kW)} \times 9.47831 \times 10^{-4}\ \text{Btu/J}}{\text{Heat Energy Input (Btu/s)}}$$

Comparison with the previous equations yields

$$\text{Efficiency} = \text{Conversion Factor (Btu/kWh)/Heat Rate (Btu/kWh)}$$

The conversion factor's role is strictly to convert the units. In the units shown it is equal to

$$\begin{aligned}\text{Conversion Factor} &= 9.47831 \times 10^{-4}\ \text{Btu/J} \times 1000\ \text{J/kW-s} \\ &\quad \times 3600\ \text{s/h} \\ &= 3412.19\ \text{Btu/kWh}\end{aligned}$$

Table 3 Heating values of some common fuels

	HHV		LHV		
Fuel	**Btu/lbm**	**Btu/gal**	**Btu/lbm**	**Btu/gal**	**LHV/HHV**
No. 2 oil	19,580	142,031	18,421	133,623	0.9408
No. 4 oil	18,890	146,476	17,804	138,055	0.9425
No. 6 oil	18,270	150,808	17,312	142,901	0.9476
Diesel fuel	19,733		18,487		0.9368
Hydrogen	61,007		51,635		0.8464
Methane	23,876		21,518		0.9012
Typical natural gas	22,615		20,450		0.9019
Propane	21,653		19,922		0.9201
Butane	21,266		19,623		0.9227
Gasoline	19,657	121,808	18,434	114,235	0.9379
Reformulated gasoline	19,545	120,103	18,304	112,377	0.9365
Methanol	11,274	73,882	10,115	66,289	0.8972
Anthracite coal	14,661		14,317		0.9765
Bituminous coal	14,100		13,600		0.9645

Source: From Based on calculations by the author and data from The Association of Energy Engineers (see Refs. 4–6).

When the power-generating equipment is used to produce electricity, the efficiency is often called the electrical efficiency.

The efficiency calculated by the above equations is based on the hope of converting the entire heating value of the fuel into useful work. It is unfortunate that this has come to be the standard because there is a further limitation imposed by the second law of thermodynamics. It tells us that nature will not allow conversion of this much chemical energy into useful work, even if the power cycle were constructed as perfectly as possible under the most ideal conditions.

THE SECOND LAW OF THERMODYNAMICS

Reversibility

Before stating the second law specifically, it is necessary to introduce the concept of reversibility. Most processes produce changes in the substances on which they operate. Reversibility has to do with how easy it would be to undo the changes that are produced in a substance through some process. Specifically, reversibility means that a change that a process makes on a substance could theoretically be undone (reversed) with only an infinitesimal change in the conditions in the process.

For example, suppose that heat is exchanged by allowing heat to flow from a hotter substance to a cooler one. For this change to be reversible, temperatures at every point where heat transfer takes place must differ only infinitesimally. Thus, if we raised the temperature profile of the cooler substance by just an infinitesimal amount, the heat could be transferred back to the hotter substance.

As another example, suppose that a gas is compressed in a cylinder by moving a piston in such a way that it reduces its volume. For the change to be reversible, the pressure the piston exerts must be only differentially greater than the pressure of the gas. In this case, if the piston pressure were reduced only infinitesimally, the gas could be expanded back to its original state, and all the work of compression would be recovered. On the other hand, if a pressure were applied that differed substantially from that of the gas in the piston, the change brought about would be irreversible, because a differential change in the applied pressure could not reverse the effects of the compression.

Formulation

The second law of thermodynamics has a certain mystique to it because unlike most physical laws, it is not a conservation law (e.g., conservation of mass, energy, momentum, etc.). Instead, it puts certain limits on what nature allows in spite of the great conservation laws. It is also not as intuitive as the other great laws. Therefore, much has been written to try to explain it. However, it can be applied rigorously based on the following three-part formulation. This formulation refers to a given substance of constant mass that experiences changes from an initial to a final state through the possible exchange of heat and work with its surroundings.

Part 1: There is a thermodynamic state function, called entropy, S, which is defined by the following equation:

$$dS = \frac{dQ_{rev}}{T} \tag{4}$$

where dS is the differential change in entropy and dQ_{rev} refers to a differential amount of heat transferred to or from a substance in a reversible manner. That is, to calculate finite entropy differences, this equation needs to be integrated along a reversible path.

Stating that entropy is a state function is not a trivial claim. What it means is that no matter what reversible path is chosen, the net change in entropy will be the same, provided that the initial and final states are the same.

Part 2: No matter how a change of state is brought about in a given substance,

$$\Delta S_{substance} + \Delta S_{surroundings} \geq 0 \tag{5}$$

The differences implied by the Δs are differences between the end state and the initial state of both the substance and surroundings. Notice that there is such a thing as a change in the entropy of the surroundings as well as the substance. It too can be calculated by Eq. 4.

Part 3: The inequality in the previous equation becomes an equality if, and only if, the process used to bring about the change is completely internally reversible and if the process also exchanges heat with the surroundings in a completely reversible manner.

Thus, this third part becomes a criterion of reversibility for a process or piece of equipment. Furthermore, the greater the inequality, the greater the irreversibility.

In general, the surroundings are considered to be at a constant ambient temperature, T_0. Thus,

$$\Delta S_{surroundings} = \frac{Q_{surroundings}}{T_0} = -\frac{Q_{substance}}{T_0} \tag{6}$$

where the two *Qs* are heat transfers to the surroundings and substance, respectively, which are the same in magnitude but opposite in sign. This equation can be combined with Eq. 5 to give

$$Q \leq T_0 \Delta S \tag{7}$$

Along with the first law for flow processes, this gives

$$W \leq T_0 \Delta S - \Delta H \tag{8}$$

In these equations, the subscripts have been dropped and all quantities refer to the substance, not the surroundings. Furthermore, although these equations were derived for changes in a substance of constant mass (closed system), as with the first law, they can also be applied to a steady flow system. In this case, the Δ*s* refer to differences in the fluxes of the thermodynamic properties between the output and inlet streams.

The implications of these two equations are very significant. They mean that in a flow process, if the inlet and outlet conditions of the various streams are fixed, there are an upper bound to the amount of heat that can be transferred to the process and an upper bound on the amount of work it can produce. The latter claim has an impact on power generation systems because it imposes a limit on the amount of power that can be produced from a given amount of fuel, regardless of its energy content. Furthermore, because the entropy change of such processes is usually negative, another implication of Eq. 7 is that a certain amount of heat must be rejected to the surroundings regardless of how well the process is designed internally.

The Carnot Engine

Some of the implications of the second law can be illustrated through the so-called Carnot engine. This is an imaginary reversible engine in which heat, Q_H, is transferred from a high-temperature source at T_H into a reversible engine, which produces work, *W*, and discards heat, Q_C, into a cold sink—possibly the environment—at T_C (Fig. 3). From the first law,

$$W = Q_H - Q_C$$

From the second law,

$$\frac{Q_C}{T_C} - \frac{Q_H}{T_H} \geq 0$$

If the Carnot efficiency is defined as the amount of work that can be produced from the high-temperature heat (i.e., W/Q_H),

$$\frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} = \frac{T_H - T_C}{T_H} \tag{9}$$

The last term on the right is of particular interest. It means that the Carnot efficiency depends on only the

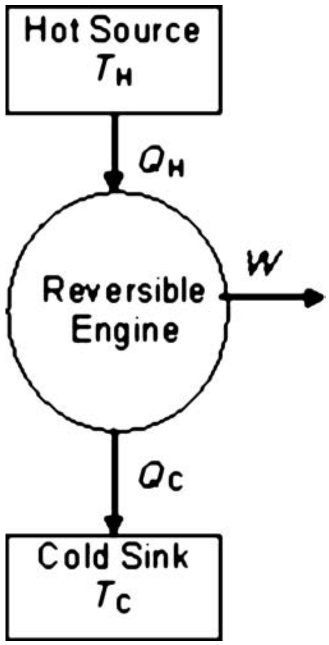


Fig. 3 The Carnot engine.

temperatures of the two heat reservoirs. All the heat in a high-temperature source can never be converted to work, even if the apparatus is perfectly reversible. The higher the temperature of the hot source relative to the cold sink, the higher the Carnot efficiency.

Ideal Work, Lost Work

The second law gives a means to measure the irreversibility of a process or piece of equipment. That, in turn, is a measure of its departure from an ideal design—one which makes changes to substances in a way that most preserves their work-producing potential. This analysis begins by imagining a perfectly reversible process and defining W_{ideal} and Q_{ideal} to be the limiting values defined by Eqs. 7 and 8 (see Fig. 4).

$$W_{\text{ideal}} = T_0 \Delta S - \Delta H \tag{10}$$

$$Q_{\text{ideal}} = T_0 \Delta S \tag{11}$$

These ideal values can be determined from the enthalpies and entropies of the various streams entering and leaving the process alone, without reference to any of the internal details of the process.

The level of irreversibility in a real process or in some part of it can be quantified through what is called lost work.[7] Because the actual work that can be obtained from any real process is always less than that what would have resulted from a completely reversible process, this leads to the following definition:

$$W_{\text{lost}} = W_{\text{ideal}} - W \tag{12}$$

By combining this with Eqs. 10 and 3, this becomes

$$W_{\text{lost}} = T_0 \Delta S - Q \tag{13}$$

A real process—or any subprocess—can be represented as shown in Fig. 5. The W_{lost} stream is not an actual work output. However, showing it in this manner is a reminder that in any real process a certain amount of the potential to do work has been lost due to the irreversibilities of the process. Furthermore, the lost work of all the subprocesses within the overall process sum up to the lost work of the process as a whole,

$$W_{\text{ideal}} = W + \sum W_{\text{lost}} \tag{14}$$

where the summation is carried out over each of the subprocesses. When analyzing a process, Eq. 13 can be applied to each piece of equipment. Then the W_{lost} for each of these can be summed according to Eq. 14. Such an analysis brings out where the greatest irreversibilities occur and therefore shows where the greatest opportunities are for improvement.

Exergy

Lost work is related to the modern concept of exergy. Lost work gives a quantitative measure of the irreversibility of a process or subprocesses within it. Furthermore, the term "lost work" seems quite descriptive. However, it is also possible to assign values—perhaps even monetary values—to streams as well as equipment. This can be done through the concept of exergy.

The exergy of a stream is the maximum amount of work that could theoretically be extracted from that stream if its temperature and pressure were reduced to that of the environment and if the concentrations of its chemical species were brought to that of the environment. It should be clear that from a power-production point of view, the economic value of a stream is related to its exergy because the production of work is usually more valuable than the production of heat alone.

Exergy analysis is based on the idea that it costs something in terms of equipment, power, adding costly streams, etc. to raise a stream's exergy—and hence its value. If what it costs to raise the value of a given stream is less than the increased value of the stream, it is a desirable thing to do.

Exergy is generally divided into four parts: physical, chemical, kinetic (having to do with the velocity of the fluid), and potential (having to do with its height above some datum level). The first two of these parts are of primary interest here. Physical exergy is the negative of the ideal amount of work that could be extracted from a stream if its temperature and pressure were reduced to ambient conditions. That is,

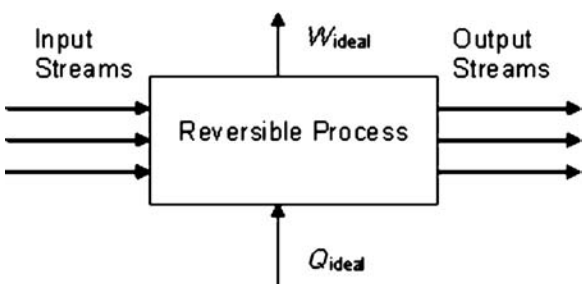


Fig. 4 Schematic representation of a reversible thermodynamic process.

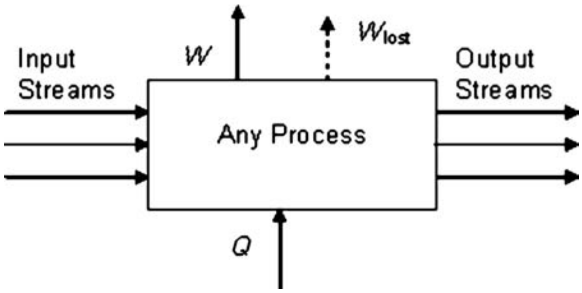


Fig. 5 Schematic representation of any process showing lost work.

$$e^{\mathrm{Ph}} = \Delta H - T_0 \Delta S \tag{15}$$

In Eq. 15 and the following equations, the Δs refer to differences between the present state of the fluid and its state at ambient conditions.

Chemical exergy, e^{Ch}, also has to be added because the exergy, say, of a fuel at ambient conditions ought to be higher than that of its eventual combustion products at those conditions because it has the potential of producing work in being transformed into its products. This can be accounted for in an equation similar to the one above, but this time involving properties of formation at ambient conditions.

$$e^{\mathrm{Ch}} = \Delta H_{\mathrm{f}} - T_0 \Delta S_{\mathrm{f}} \tag{16}$$

Exergy is not quite a thermodynamic state function because it depends on environmental temperature. Also, unlike enthalpy or internal energy, it is not conserved in a process. Nevertheless, it is a useful measure of the power-producing value of a process stream. In any real processes, every process step results in a net loss of exergy.

The relationship between exergy and lost work is very straightforward—namely,

$$\text{Lost Work} = \sum \text{Exergies of incoming streams} - \sum \text{Exergies of outgoing streams} \tag{17}$$

where the summations are carried out over all incoming and outgoing streams, respectively.

Reversibility Revisited

The preceding analysis shows how the second law can be used to provide a measure of the irreversibility of a process or part of a process. Every irreversibility causes a loss in the potential to convert energy into work. In this era of diminishing supplies of cheap fuels and increasing demands for electric power, this is an important consideration. Furthermore, irrersibilities in one part of a process reduce the amount of work that can be produced—or else increase the amount of work that is needed—in the rest of the process, even though the source of the irreversibility may be far removed from the place in the process where work is produced or consumed.

The following is a list of some process steps that introduce irreversibility into processes, including power cycles:

- Heat transfer across a nonzero temperature difference
- Mixing of fluids of differing pressures, temperatures, or compositions
- Pressure drops that do not recover work
- Chemical reactions that do not take place at equilibrium conditions
- Flashing of liquids when pressure is reduced
- Two-phase contact and mass transfer between fluids whose various species have concentrations that differ from their equilibrium values
- Temperature and pressure shocks when fluids enter equipment at different temperatures or pressures from those that are present within the equipment

The most thermodynamically efficient designs—those that most preserve the work-producing potential of the fluids in the process—are those that strive for the greatest reversibilities in the various pieces of process equipment. However, there is another side to this question. The more reversible a piece of equipment is, the greater its capital cost is apt to be. Thus, in the practical design of equipment there is often a trade-off between reversibility and capital cost. Processes need to be evaluated not only for their thermodynamic efficiency, but also for their thermoeconomic effectiveness.

CALCULATION OF THERMODYNAMIC PROPERTIES

The Rigorous Equations

Process equipment cannot be designed or evaluated apart from the ability to estimate the thermodynamic properties of the entering and exiting streams. The following are rigorous equations for the calculation of some of the important thermodynamic properties of pure or constant-composition substances[8]:

$$\mathrm{d}H = C_p \mathrm{d}T + \left[V - T\left(\frac{\partial V}{\partial T}\right)_{\mathrm{P}}\right]\mathrm{d}P \tag{18}$$

$$\mathrm{d}S = \left(\frac{C_p}{T}\right)\mathrm{d}T - \left(\frac{\partial V}{\partial T}\right)_{\mathrm{P}}\mathrm{d}P \tag{19}$$

$$G = H - TS \tag{20}$$

C_p is the heat capacity at constant pressure, defined by

$$C_p = \left(\frac{\partial H}{\partial T}\right)_{\mathrm{P}} \tag{21}$$

G is called the Gibbs free energy or Gibbs function and has importance in equilibrium calculations. Eq. 20 can be considered its definition. By differentiation and comparison with the other two, it can be shown that

$$\mathrm{d}G = V\mathrm{d}P - S\mathrm{d}T \tag{22}$$

These are differential equations that emphasize that these properties need reference values from which differences can be calculated. To obtain values for the property differences for real substances, the

Ther–Und

pressure–volume–temperature (PVT) behavior of the fluid needs to be inserted into the equations, which would then be integrated to produce the final results.

The Ideal Gas

One way of inserting PVT behavior is through equations of state. One of the simplest models for gases is that of the ideal gas, whose equation of state on a molar basis is:

$$PV = RT \tag{23}$$

This equation reflects the fact that the molecules in the gas do not interact in any way, although they may have very complex energy interactions within a given molecule.

When this equation is inserted into the rigorous equations, one obtains

$$dH^* = C_p^* dT \tag{24}$$

$$dS^* = \left(\frac{C_p^*}{T}\right) dT + \left(\frac{R}{P}\right) dP \tag{25}$$

$$dG^* = RT d \ln P - S^* dT \tag{26}$$

where the asterisk designates the properties as those of an ideal gas.

The integrated forms of Eqs. 24 and 25 are

$$\Delta H^*(P2, T2; P1, T1) = \int_{T1}^{T2} C_p^* dT \tag{27}$$

$$\Delta S^*(P2,T2; P1,T1) = \int_{T1}^{T2} \left(\frac{C_p^*}{T}\right) dT + R \ln\left(\frac{P2}{P1}\right) \tag{28}$$

The heat capacity has been left behind the integral sign because even for most ideal gases, it varies with temperature. However, neither ideal gas heat capacities nor enthalpies vary with pressure, but ideal gas entropies do. Ideal gas heat capacity data is often presented in the form of an analytical equation. One common form used by NIST[9] is

$$C_p^* = A + BT + CT^2 + DT^3 + \frac{E}{T^2} \tag{29}$$

NIST tabulates the constants (A, B, etc.) for many compounds.

Real Fluids

Real fluids may be represented by the following general equation of state:

$$V = Z(P,T)RT/P \tag{30}$$

where Z is the compressibility factor, which can be a function of both P and T. It is common to report PVT data in the form of the compressibility factor. For an ideal gas, $Z=1$. Thus, the compressibility factor is a measure of the deviation from ideal gas behavior.

When Eq. 30 is substituted in Eqs. 18 and 19, one obtains

$$dH = C_p dT - \left[\frac{RT^2}{P}\left(\frac{\partial Z}{\partial T}\right)_P\right] dP \tag{31}$$

$$dS = \left(\frac{C_p}{T}\right) dT - R\left(Z + T\frac{\partial Z}{\partial T}\right)\frac{dP}{P} \tag{32}$$

One consequence of these equations is that thermodynamic properties of real fluids can be calculated from their ideal gas–heat capacity (as a function of temperature) and their PVT behavior. This follows from the fact that to get from one state to another, these equations can first be integrated to zero pressure (which requires only PVT data), then integrated at zero pressure to the final temperature (which requires only ideal gas heat capacity data), and then integrated back to the end-point pressure.

Several approaches have been used to apply these equations for practical calculations. A first approach is to take known ideal gas-heat capacity and PVT data, and use them directly in these equations. Such data are not often available, and even when they are, this is a tedious process. Nevertheless, it has been done for a number of well-studied chemical species, most notably water. This is the basis of the steam tables, which have been put into analytical form for easy computer calculations.[10]

A second approach is through generalized correlations. An early attempt to describe the PVT behavior of many fluids was begun by Pitzer[11] through the use of his so-called acentric factor. His correlations have been extended so that they can be used to predict the more important thermodynamic properties of real no-polar fluids in both vapor and liquid phases.[12]

A third approach is to insert an equation of state into Eqs. 31 and 32 so that they may be integrated analytically. One excellent equation of state for low-density gases is the virial equation

$$Z = 1 + \frac{B(T)}{V} + \frac{C(T)}{V^2} + \frac{D(T)}{V^3} + \cdots. \tag{33}$$

where B, C, D, etc. are functions of temperature only and are called the second, third, fourth, etc. virial coefficients. There is a great deal of data for the second virial coefficients.[13] Where data are not available, correlations have been developed to estimate them.[14,15] Third virial coefficients are usually not available, but prediction methods have been proposed.[16] Very little information is available for higher-order coefficients. Other equations of state have also enjoyed success in predicting properties of real fluids, both in the vapor and liquid phases. Among the more successful are the Soave-Redlich-Kwong[17] and Peng-Robinson[18] equations.

Solutions

The properties of solutions—whether they are gas or liquid mixtures—are not merely the mole fraction averages of the properties of the pure components. For any thermodynamic property, M, the exact equation for the properties of mixtures is

$$M(P,T,\bar{x}) = \sum_i x_i \overline{M_i}(P,T,\bar{x}) \tag{34}$$

where $\overline{M_i}$ is called the "partial molal property of i" and $\bar{x}$ is a vector containing all of the compositions of the various constituents of the solution. The summation is carried out over all constituents. $\overline{M_i}$ is the property of constituent i at pressure, P, and temperature, T, as it exists in solution of composition $\bar{x}$. In general, this is different from its value as a pure substance.

The most general equation through which partial molal properties may be evaluated is

$$\overline{M_i} = \left(\frac{\partial(nM)}{\partial n_i}\right)_{P,T,n_j} \tag{35}$$

where n is the total number of moles and n_i is the number of moles of specie "i".

The partial derivative implies that the extensive property, nM, is differentiated with respect to the particular specie of interest, i, with pressure, temperature, and all of the other constituent amounts held constant. If properties of the various possible mixtures are known or can be approximated analytically, this equation provides a means to calculate their partial molal properties.

An important approximation for solutions is that of so-called "ideal solutions," whose properties are calculated as follows,

$$V^{\text{id}} = \sum_i x_i V_i \tag{36}$$

$$H^{\text{id}} = \sum_i x_i H_i \tag{37}$$

$$S^{\text{id}} = \sum_i x_i S_i - R \sum_i x_i \ln x_i \tag{38}$$

$$G^{\text{id}} = \sum_i x_i G_i + RT \sum_i x_i \ln x_i \tag{39}$$

where the superscript, "id," designates these as the properties for ideal solutions. All ideal gases are ideal solutions. The ideal-solution approximation is useful in some situations, but most liquids depart from it significantly. Notice also that properties involving entropy have a mixing effect (second terms on the right side of Eqs. 38 and 39) even if they are ideal solutions. This reflects the fact that the entropy of a solution is always higher than the combined entropies of its constituents.

The true properties of solutions can be calculated from their PVT behavior. For humid air, for example, the mixing effect is often neglected. However, one recent study done by *M*. Conde Company[19] includes virial coefficient information, including information for mixtures, sufficient to produce very accurate psychrometric properties of air–water mixtures.

Phase Equilibrium

Phase equilibrium can also be predicted from thermodynamic properties. The way this is done goes beyond the scope of this entry. However, several ideas can be stated here. One aspect of phase equilibria—namely pure-component vapor pressures—has been widely measured and can be used as a starting point for multicomponent phase equilibrium estimates. Vapor pressure data is often correlated by the Antoine equation,

$$\log_{10}(P^{\text{V}}) = A - \left(\frac{B}{T + C}\right) \tag{40}$$

where A, B, and C are constants. Such constants for many compounds also appear in the NIST collection.[20]

One special case of interest in energy calculations is the two-phase equilibrium between (1) pure liquid water and (2) a vapor phase one of whose components is water vapor. If the vapor phase is considered to be an ideal gas (and thus an ideal mixture), and if certain other assumptions are made, the composition of water in the vapor phase can be estimated by

$$y_{H_2O} = \frac{P^{\text{V}}(T)}{P} \tag{41}$$

where the vapor pressure can be calculated from the Antoine equation. This is a useful approximation that can be refined if PVT data for the vapor mixture are available.

CONCLUSIONS

This entry attempts to give an introduction to some of the basic principles of thermodynamics and how they relate to various energy engineering applications. The two main laws of thermodynamics govern the performance of any energy-conversion process. Their application requires the ability to estimate the thermodynamic properties of the various process streams. Such estimates are also based on the formalism that thermodynamics provides. Thus, thermodynamics provides the basic framework within which the design and evaluation of all energy-conversion equipment and processes must be performed.

ACKNOWLEDGMENT

Support for this work by VAST Power Systems, Elkhart, Indiana, is gratefully acknowledged.

REFERENCES

1. The ideal gas state is difficult to describe briefly. It is a fluid with the same heat capacity as the real fluid's heat capacity at very low pressure. However, unlike the real fluid, it obeys the ideal gas law. For low-pressure calculations, the difference between the ideal gas state and the real fluid at low pressure is not important.
2. http://webbook.nist.gov/chemistry/form-ser.html (accessed July 2005).
3. ISO conditions for air also include a relative humidity of 60%.
4. Notes from short course: Capehart, B.L.; Salas, C.E. *The Theory and Practice of Distributed Generation and On-Site CHP Systems*; The Association of Energy Engineers: Atlanta, GA, June 2003.
5. Bosul, Ulf. *Well-to-Wheel Studies, Heating Values and the Energy Conversation Principle*, http://www.efce.com/reports/E10.pdf (accessed July 2005).
6. http://www.chpcentermw.org/pdfs/toolkit/7c_rules_thumb.pdf (accessed July 2005).
7. Van Ness, H.C.; Abbott, M.M. Thermodynamics. In *Perry's Chemical Engineers' Handbook,* 7th Ed.; Perry, R.H., Green, D.W., Eds.; McGraw Hill: New York, 1997; 4-1–4-36; However, note that because of a different sign convention their formulas for ideal work have the opposite signs from those in this article.
8. The equations can be found in many thermodynamics text books. A concise useful summary is found in: Van Ness, H.C.; Abbott, M.M. In *Perry's Chemical Engineers' Handbook,* 7th Ed.; Perry, R.H., Green, D.W., Eds.; McGraw Hill: New York, 1997; 4-1–4-36.
9. http://webbook.nist.gov/chemistry/form-ser.html (accessed July 2005).
10. http://www.cheresources.com/iapwsif971.pdf (accessed July 2005). A free downloadable Excel add-in which embodies the equations is also available on this Web site.
11. Pitzer, K.S. *Thermodynamics*, 3rd Ed.; Mc-Graw Hill: New York, 1995; App. 3.
12. Lee, B.I.; Kessler, M.G. A generalized thermodynamic correlation based on three-parameter corresponding states. AIChE J. **1975**, *21* (3), 510–527.
13. Daubert, T.E.; Danner, R.P. *Physical and Thermodynamic Properties of Pure Chemicals Data Compilation*; Taylor & Francis: Bristol, PA, 1994.
14. Tsonopoulos, C. An empirical correlation of second virial coefficients. AIChE J. **1974**, *20* (2), 263–272.
15. Liley, P.E.; Thomson, G.H.; Daubert, T.E.; Buck, E. Physical and chemical data. In *Perry's Chemical Engineers' Handbook,* 7th Ed.; Perry, R.H., Green, D.W., Eds.; McGraw Hill: New York, 1997; 2-355–2-358.
16. De Santis, R.; Grande, B. An equation for predicting third virial coefficients of nonpolar gases. AIChE J. **1979**, *25* (6), 931–938.
17. Soave, G. Equilibrium constants from a modified Redlich-Kwong equation of state. Chem. Eng. Sci. **1972**, *27*, 1197–1203.
18. Peng, D.-Y.; Robinson, B.B. A new two-constant equation of state. Ind. Eng. Chem. Fund. **1976**, *15*, 59–64.
19. http://www.mrc-eng.com/Downloads/Moist%20Air%20Props%20English.pdf (accessed July 2005).
20. http://webbook.nist.gov/chemistry/form-ser.html (accessed July 2005).

Tradable Certificates for Energy Savings

Silvia Rezessy
Energy Efficiency Advisory, REEEP International Secretariat, Vienna International Centre, Austria, Environmental Sciences and Policy Department, Central European University, Nador, Hungary

Paolo Bertoldi
European Commission, Directorate General JRC, Ispra (VA), Italy

Abstract

Recently tradable certificates for energy savings have attracted the attention of policy makers as a tool to stimulate energy efficiency investments and deliver energy savings. While such schemes have been introduced in different forms in Italy, France and Great Britain and considered in other European countries, there is an ongoing debate over their effectiveness and applicability. The paper describes the concept and main elements of schemes that involve tradable certificates for energy savings (TCES) and how these have been put into practice in Italy, France and Great Britain. The entry discusses some key design and operational features of TCES schemes such as scope, additionality rules, measurement and verification. The implications of different certificate trading rules and how these can affect the actual structure of the certificate market are also explored.

INTRODUCTION

Market-based instruments (MBIs [MBIs are public policies that make use of market mechanisms with transferable property rights to distribute the burden from a policy. We recognize the difference between policy instruments that are well positioned to harness market forces to achieve a certain policy goal (such as renewable energy quotas or renewable portfolio standards [RPS]) and the market instruments (namely, carbon allowances, and green and white certificates), the latter being a much narrower concept representing just a tradable commodity. This differentiation is not so important in the context of the present entry, and in the text, we refer to complex policy tools/portfolios that include trading of financial commodities (such as certificates or allowances) as that aim at MBIs.) that aim at bringing sustainability to the energy sector have been implemented to promote electricity from renewable energy sources and to cut harmful emissions. Quota systems (also known as RPS) coupled with tradable green certificate (TGC) schemes have been developed and tested in several European countries to foster market-driven penetration of renewable energy sources. Another well-known and widely analyzed type of MBI is the tradable emission allowance.

To stimulate energy-efficiency investments and to achieve national energy savings targets, the attention of policy-makers in Europe has recently been attracted by the possibility of introducing energy savings obligations on certain types of market players coupled with tradable certificates for energy savings (Tradable Certificates for Energy Savings [TCES], or white certificates). Such schemes have been introduced in different forms in Italy, Great Britain, and France, and are being considered in other European countries.

This entry describes the concept and main elements of a TCES scheme and compares and analyses how these have been put into practice in Italy, Great Britain, and France. The entry builds on a sequence of publications by the authors, giving specific details on the still fairly short implementation track record of TCES schemes, as well as qualitative comparison of the TCES scheme with other policy tools promoting energy efficiency and the possibilities and dangers associated with attempts to integrate existing MBIs into the energy sector.[2–5]

SCHEMES WITH TRADABLE CERTIFICATES FOR ENERGY SAVINGS

Energy efficiency is a well-established option to decouple economic growth from the increase in energy consumption and thus reduce greenhouse gas (GHG) emissions by cutting the amount of energy required for a particular amount of end-use energy service. Apart from being a sound part of the environmental and climate change agenda, increased energy efficiency can contribute to meeting widely accepted goals of energy policy such as improved security of supply, economic efficiency, and increased business competitiveness coupled with job creation and improved consumer welfare. Alongside environmental and economic sustainability, the other main driver of energy policy in the European Union (EU) is to restructure electricity and gas markets. Many energy-efficiency advocates and policy-makers have

Keywords: Energy saving obligations; Tradable certificates for energy savings; Energy suppliers.

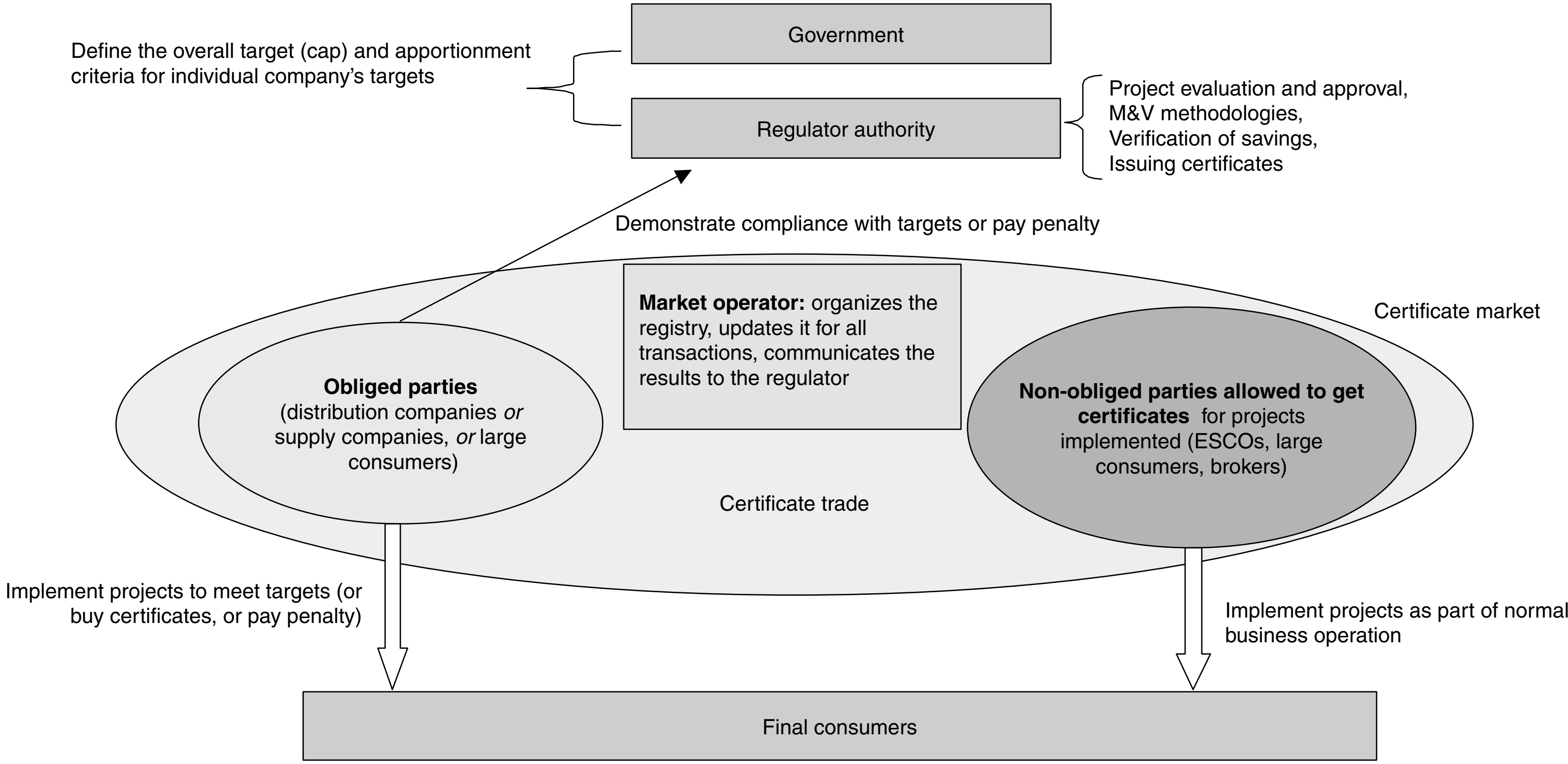


Fig. 1 A policy portfolio with mandatory savings targets and white certificates, and a summary of roles of actors and relationships among them.

called for legislation introducing energy efficiency and energy services as natural complements to the electricity and gas market liberalization. Otherwise, market failures in the energy sector would lead to lower levels of investment in energy efficiency than is socially optimal, with the final outcome being additional cost to the economy due to an imbalance between the supply side and demand side in the energy sector.

A possible market-based policy portfolio could comprise energy-savings quotas for some category of energy operators (distributors, suppliers, consumers, etc.) coupled with a trading system for energy-efficiency measures resulting in energy savings. The savings would be verified by the regulator and certified by means of the so-called white certificates (tradable certificates for energy savings). In Italy certificates are called Energy Efficiency Titles, while in France they are referred to as Certificates of Energy Savings.

A TCES portfolio involves the following basic elements[2,3,5,7,12,13]:

- The creation and framing of the demand. Tradable certificates represent a meaningful option only if there is interest in buying/selling them.
- The tradable instrument (certificate), representing the savings, conferring property rights to the holder, and providing the rules for trading.
- Institutional infrastructure and processes to support the scheme and creation of the market. An aspect that is often overlooked is that markets do not function in a vacuum. These activities include, for instance, measurement and verification (M&V), evaluation methods and rules for issuing certificates, a data management and certificate tracking system, and a registry.
- Cost-recovery mechanisms, in some cases.

Fig. 1 illustrates the roles and summarizes the responsibilities of different categories of actors involved in the design and operation of a TCES policy portfolio.

GENERAL CHARACTERISTICS OF ENERGY-SAVING OBLIGATIONS AND TRADING SYSTEMS IN EUROPE

Variations of the policy mix described above have been introduced in Italy, Great Britain, and (since July 2006) France. In the Flemish region of Belgium, savings obligations are imposed on electricity distributors without any trading option (certificate or, as in Great Britain, obligation trading). The first operational scheme in the world with a white-certificate trading element has been introduced in New South Wales (NSW), Australia. It is, however, a GHG trading system that has an end-use energy efficiency element and is left out of the present discussion.

Box 1. Tradable certificate for energy savings (white certificate)

A White certificate is an instrument issued by an authority or an authorized body providing a guarantee that a certain amount of energy savings has been achieved. Each certificate is a unique and traceable commodity that carries a property right over a certain amount of additional savings and guarantees that the benefit of these savings has not been accounted for elsewhere.

In Italy, command-and-control measures (energy savings targets in primary energy consumption for electricity and gas grid distribution companies with more than 100,000 customers as of the end of 2001) are combined with market instruments (tradable certificates for energy savings issued to distributors and energy service companies) as well as with elements of tariff regulation (a cost-recovery mechanism via electricity and gas tariffs and multiple-driver tariff (MDT) schemes (MDTs essentially constitute tariff regulation schemes linking the evolution over time of allowed revenue with cost drivers such as the number of customers, grid lengths, and energy sales.[17]) to reduce the disincentives for regulated electricity and natural gas companies to promote end-use energy efficiency among their customers) or dedicated funds in some circumstances. Over the 5 years of the current phase of the scheme, 3 million tons of oil-equivalent (Mtoe) cumulative primary energy savings are projected to be realized, of which 1.6 Mtoe is by electricity distributors and 1.3 Mtoe is by natural gas distributors. At least half of the target set for each single year is to be achieved via a reduction of electricity and gas end-use consumption (referred to as the 50% constraint, to which each distributor is subject). The remaining share can be achieved via primary energy savings in all the other end-use sectors. Energy-saving projects contribute to the achievement of targets for up to 5 years (with only some exceptions). Only savings that are additional to spontaneous market trends and legislative requirements are considered. After a long process of designing and elaborating elements, the Italian scheme finally became operational in January 2005.[12,14–16]

In Great Britain, the Energy Efficiency Commitment (EEC) runs in 3-year cycles from 2002 to 2011. It replaced the Energy Efficiency Standards of Performance (EESOP), running from 1994 until 2002, which established the principle of pooled spending on energy efficiency for domestic consumers. The EEC-1 program required that all gas and electricity suppliers with 15,000 or more domestic customers deliver a certain quantity of fuel standardized energy benefits by encouraging or assisting customers to take energy-efficiency measures in their homes. The overall savings target was 62 fuel standardized TWh (Energy savings are discounted over the lifetime of the measure and then standardized according to the carbon content of the fuel saved.) (lifetime discounted), and the total delivered savings reached 86.8 TWh.[10] In EEC-2 (2005–2008), the threshold for obligation was increased to 50,000 domestic customers. The target has been increased to 130 TWh; however, due to the carrying over of savings from EEC-1, in 2005 more than a quarter of this target has already been achieved. Suppliers must achieve at least half of their energy savings in households on income-related benefits and tax credits. Projects can be related to electricity, gas, coal, oil, and Liquefied Petroleum Gas (LPG). Suppliers are not limited to assisting their own customers only and can achieve improvements in relation to any domestic consumers in the Great Britain. Carbon-benefits estimations take into account the rebound effect—the likely proportion of the investment to be taken up by improved comfort—by adjusting the benefits to comfort factors. In addition, dead-weight factors are considered to account for the effect of investments that would be made anyway. At present, certificate trading is not a feature of the scheme in Great Britain.

The French system, introduced in 2006, envisages that all electricity, gas, LPG, (Domestic fuel excludes transport usages.) oil, cooling, and heating fuel for stationary applications suppliers that supply over 0.4 TWh/year will have to meet a target of energy savings. It excludes plants under the EU Emission Trading System and fuel substitution between fossil fuels, as well as energy savings resulting only from measures implemented just to comply with current legislation. Apart from these, no additional restrictions on compliance are foreseen. To meet the obligation, savings can be made in any sector and with any energy source or carrier. The following details related to the obligation parameters are known at the time of finalizing this entry: the total target for the first 3 years (2006–2008) will be 54 TWh (in final energy), cumulated over the life of the energy-efficiency actions with a 4% discount rate. The target apportionment is a two-step procedure: first, the total obligation is divided among different energies; then the obligation for a particular energy is apportioned among respective energy suppliers included in the system. The expected cost of action is below 20 Euro/MWh.[1]

Energy-efficiency obligations without certificate trading are also in place in the Flemish region of Belgium. Regional utility obligations were introduced in 2003 and are imposed on the electricity distributors. Currently, 16 electricity distributors are covered by the obligation. The annual target is 0.58 TWh; eligible actions refer to residential and nonenergy-intensive industry and service, and can involve saving fuel from any sources. Separate targets are set for low-voltage clients (<1 kV, mainly residential) and high-voltage clients (>1 kV). For the low-voltage clients, the target is 10.5% of electricity supplied over the 6 years from 2003 to 2008, and for high-voltage users (>1 kV), the target is 1% per annum for each over the same period. The reason for the higher than 1% per annum target for the low-voltage users is because of the Flemish Parliament's decision to provide free vouchers for the head of every family in 2004 and 2005, which can be exchanged via the electricity distributor for either an energy-saving Compact Fluorescent Lamps (CFL), a low-flow shower head, or an energy meter. In 2006 and 2007, it is planned that the other family members will receive a voucher for an energy-efficient light bulb.[9] There is a discussion going on in the Netherlands about the introduction of a white-certificate scheme.

ENERGY-SAVING OBLIGATIONS AND TRADING SYSTEMS: EUROPEAN EXPERIENCES TO DATE

Below, details on and first experiences with the following parameters of the existing European schemes with energy-saving obligations and energy-saving trading elements are reviewed: (a) eligible projects allowed; (b) institutional infrastructure and processes to support the scheme; and (c) certificate delineation, trading rules, and tools to stabilize the market. A comprehensive discussion of these and other design and operational features is available in Bertoldi and Rezessy.[3]

Eligible Projects

In Italy, projects in all end-use sectors are eligible. At least half of the target set for each single year should be achieved by reduction of the supplied energy vector—i.e., electricity and gas uses, a.k.a. the 50% constraint.[12] The remaining share can be achieved via primary energy savings in all of the other end-use sectors. There is an illustrative list of eligible projects. Energy savings projects contribute to the achievement of targets for up to 5 years (with only some exceptions). Early experience in Italy shows that a significant share of savings certified in autumn 2005 is coming from cogeneration, district heating, and public lighting projects. There are numerous submissions for certification of projects following the deemed savings verification method, which has very minor data requirements (see more details later in this entry). A surplus of certificates and banking of certificates are expected in Italy because the regulator has to evaluate and certify the savings from eligible projects starting from 2001, when the decrees were passed.

In Great Britain, only activities concerning domestic users are eligible. At least 50% of the energy savings must be targeted to customers who receive income-related benefits or tax credits (a.k.a. the priority group). Projects can be related to electricity, gas, coal, oil, and LPG. Suppliers can achieve improvements in relation to any domestic consumers in the United Kingdom. A nonexclusive list of measures is included within the illustrative mix for EEC 2005–2008. Measures that are related to the reduction of energy vectors other than the one supplied by the obliged party are allowed. Experience from EEC-1 in Great Britain shows that a significant share (56%) of the 86.8 TWh of savings delivered in the period 2002–2005 came from building insulation (wall and loft). CFLs accounted for a quarter of the savings achieved, followed by appliances (11%) and heating measures (9%).[10] CFLs accounted for the largest number of projects undertaken (almost 40 million measures related to CFL installation in EEC-1), followed by wet and cold appliances.[8] All but two suppliers—which went into administration and administrative receivership—achieved their targets; six suppliers exceeded their targets in EEC-1 and carried out their additional savings to EEC-2. Suppliers can receive a 50% uplift on the savings of energy-efficiency measures that are promoted through energy service activities. This uplift is limited to 10% of the overall activity. Of the six major suppliers with an EEC target, three submitted schemes that would take them over the 10% threshold if take up had been as forecasted. In reality, the energy services uplift was only 3.6% of all insulation activity. There is uplift on innovative technologies as well.

Apart from plants under the EU Emission Trading System (ETS) directive, fuel substitution between fossil fuels, and projects resulting just from measures implemented only to conform to current legislation, no other restrictions on compliance are foreseen in the French scheme. Any economic actor can implement projects and get savings certified as long as savings are above 3 GWh over the lifetime of a project. Actions must be additional relative to their usual activity, and there is a possibility to pool savings from similar actions to reach the threshold. All energies (including fuel) and all of the sectors (including transports and excluding installations covered by ETS) are eligible. Certification of projects implemented by bodies, which do not have savings obligation, is allowed, but only after considering the impact of a project on business turnover. If impact on business turnover is identified, certification of savings is allowed only for innovative products and services. “Innovative” product in this discourse means that efficiency is at least 20% higher compared with standard equipment and market share is below 5%.

Institutional Infrastructure and Processes to Support the Scheme

A sound institutional structure is needed for a white-certificate system to function, including administrative bodies to manage the system, as well as processes such as verification, certification and market operation, transaction registry, and detection and penalization of noncompliance.

Under the EEC in Great Britain, the regulator Office of Gas and Electricity Markets (OFGEM) manages project evaluation and approval, verifies savings, and manages the data. In Italy, the regulator Authorita per l’energia elettrica e il gas (AEEG) implements the scheme; the marketplace is organized and managed by the electricity market operator Gestore del Mercato Elettrico (GME) according to rules and criteria approved by AEEG. GME issues and registers certificates upon specific request by AEEG, organizes market sessions, and registers bilateral over-the-counter contracts according to rules set by AEEG.[12] In France, certificates are issued by the Ministry of Industry, and the French Agency for Environment and Energy Management and Association Technique pour l’Environnement et l’Energie (ATEE) are in charge of the definition of standardized actions.

To determine the energy savings resulting from an energy-efficiency activity, the eventual energy consumption has to be compared with a baseline (reference situation) without additional saving efforts. The choice of the reference scenario—in terms of reference consumption and conditions—raises some challenges, such as determining the relevant system boundary, minimizing the risk of producing leakage, establishing the practicality and cost effectiveness of a baseline methodology, and treating no-regret measures in the baseline determination (We are indebted to Ole Langniss for these comments). Additionality refers to certification of genuine and durable increases in the level of energy efficiency beyond what would have occurred in the absence of the energy-efficiency intervention—for instance, due only to technical and market development trends and policies in place. While in practice, projects tend to have a mix of public and private benefits, costs of disaggregating these benefits and precisely accounting for the exact share of no-regret measures in a larger action may be prohibitively high.

One way of overcoming this problem would be to place an objectively defined discount factor on investments, which accounts for these private benefits. One possibility is to use minimum efficiency requirements or current sale-weighted average efficiency levels. Furthermore, the electricity price and the effects of the EU ETS and other policies in place (such as taxation or standards) may also be accounted for in the baseline to ensure genuine additional savings. In Great Britain, a discount factor of 3.5% over the lifetime of the measure is applied, while in France, the discount factor is 4%. However, in both the British and French schemes, the savings are cumulated over the lifetime of equipment, and the discount factor refers to actualizing the annual savings for different measures with different life spans.

In Great Britain, saving estimations take into account the likely proportion of the investment to be taken up by improved comfort (comfort factors adjustment of carbon benefits; see earlier discussion) as well as dead-weight factors to account for the effect of investments that would be made anyway.

In Great Britain, the Department for Environment, Food, and Rural Affairs (DEFRA) requires suppliers to demonstrate additionality. Concerns have been raised that energy suppliers can claim toward their EEC target the total energy savings that flow from a partnership project regardless of the actual financial contribution made by the supplier.

In Italy, savings have to go over and above spontaneous market trends or legislative requirements[14,16]; the business-as-usual trend will be adjusted with time. The nature of the additionality check differs for project types (e.g., installation of efficient equipment may be evaluated on the basis of difference with the national average installed or with what is offered in shops). For projects that are based on the deemed savings and engineering verification approach (see details below), a case-by-case additionality check is performed by the regulator.

The Italian scheme uses three valuation (M&V) approaches: (a) a deemed savings approach with default factors for free riding, delivery mechanism, and persistence; (b) an engineering approach; and (c) a third approach based on monitoring plans whereby energy savings are inferred through the measurement of energy use. In the latter case, all monitoring plans must be submitted for preapproval to the regulatory authority AEEG, and they must conform to predetermined criteria (e.g., sample size, criteria to choose the measurement technology).[14,16] In practice, most of the projects submitted to date have been of the deemed saving variety. There is ex-post verification and certification of actual energy savings achieved on a yearly basis (e.g., in the case of Combined Heat and Power [CHP], the plant operator has to prove that the plant has run a certain number of hours, etc).[11] In principle, the metering approach is a more accurate guarantee of energy saved than the standard factors approach (the latter cannot verify details such as location and operating hours of installed CFLs), but in practice it can be difficult to identify the actual savings (e.g., in households, there is only one meter for all electricity usage, which increases each year due to growth in appliances and can fluctuate with changing household numbers, lifestyle, weather, etc). It may be reasonable for large installations or projects, but it may result in high monitoring costs for smaller projects.[14,16]

In Great Britain, the savings of a project are calculated and set when a project is submitted based on a standardized estimate taking into consideration the technology used, weighted for fuel type and discounted over the lifetime of the measure. There is limited ex-post verification of the energy savings carried out by the government, although this work would not affect the way energy savings are accredited in the current scheme; the monitoring work affects the energy savings accredited in future schemes.

Energy savings can be determined by metering or estimating energy consumption ahead of time and comparing this with consumption after the implementation of one or more energy-efficiency improvement measures, adjusting for external factors such as occupancy levels and level of production. Certificates, therefore, can be issued either ex-post (representing the energy saved over a certain period) or ex-ante (representing the estimation of the energy to be saved over a certain period). With regard to ex-post certification, there are different options: the saved energy resulting from an energy-efficiency measure could be measured at the end of a predetermined period (e.g., after 1 year) or over the lifetime of the project (which has to be assessed accurately). The latter option will make the system more comparable to a green certificate. The certificate has a unique time of issue attached to it; it indicates the period over which energy has been saved, the location where energy has been saved, and who is saving the energy (initial owner of the certificate). Ex-post certification, however, will probably increase validation efforts and verification costs. Alternatively, for projects that can be monitored through a standard savings approach, certificates can be granted in advance (ex-ante) of the actual energy savings delivery. This will mitigate liquidity constraints of project implementers and allow them to finance new projects. If underperformance is detected at the end of the lifetime of the measure, the underperforming project owner should be asked to cover the shortage with certificates purchased on the spot market.

Depending on the design of the scheme, the role of the regulator may or may not include the issues of certificates and the verification of savings. For instance, third parties may be licensed to evaluate and approve projects, verify savings, and issue certificates. Then the role of the regulator would be to accredit third parties and audit their performance. With the cost of compliance being one of the major issues raised about the implementation of white-certificate schemes, this potentially reduces the overall cost. It is not as crucial which body issues the certificates, provided that these are based on verified data, which can come from the energy regulator (as is the case in Italy) or from a certified verifier.

Certificate Delineation, Trading Rules, and Tools to Stabilize the Market

The certificate is an instrument that provides a guarantee that savings have been achieved. Each certificate should be unique, traceable, and at any time have a single owner. A certificate needs to be a well-defined commodity that carries a property right over a certain amount of additional savings and guarantees that the benefit of these savings has not been accounted for elsewhere. Property rights must be clear and legally secured, as it is unlikely that trades will occur if either party is unsure of ownership.[6]

Minimum project size may be applied for certification of savings to reduce transaction costs and encourage pooling of projects.[12] The size of a certificate also has important implication on the number of parties that can offer certificates for sale (unless other restrictions apply). In Italy, certificates are expressed in primary energy saved, and the unit is 1 ton of oil equivalent (toe). In France, certification is allowed only above a threshold of 3 GWh of savings over the lifetime of a project.[1]

The validity and any associated intertemporal flexibility embodied by banking and borrowing rules, the rules for ownership transfer, the length of the compliance period, and the expectations of market actors about policy stability and continuity will all influence the market for white certificates. A long certificate lifetime and banking increase the elasticity and flexibility of demand in the long term. To mitigate the uncertainties about the achievement of the quantified policy target within the prespecified time frame, banking for obliged parties may be allowed only after they achieve their own targets. In Italy, certificates are valid for up to 5 years, with a few exceptions.[12] In Great Britain, suppliers can carry over to EEC-2 all of their excess savings from measures implemented under EEC (this refers to measures rather than savings). In France, it has been proposed that the certificates' validity be at least 10 years. Borrowing is discouraged, because it makes the attainment of a target uncertain and is against the ex-post logic of the white-certificate scheme as applied in Italy, for instance.

Rules defining trading parties are also important for market liquidity. Provided that administrative and monitoring costs are not disproportionate, as many parties should be allowed in the scheme as possible, because this enhances the prospects of diversity in marginal abatement costs and lowers the risks of excessive market power.[13] Parties that may be allowed to receive and sell certificates include obliged actors, exempt actors, Energy Service Companies (ESCOs), consumers, market intermediaries, Non Governmental Organizations (NGOs), and even manufacturers of appliances. A key benefit of allowing many parties into the scheme is that new entrants may have the incentive to innovate and deliver energy-efficiency solutions that have a lower marginal cost.

In Italy, certificates are issued by the electricity market operator upon request of the regulator AEEG to all distributors and their controlled companies and to ESCOs. Certificates are tradable via bilateral contracts or on a spot market organized and administered according to rules set out jointly by AEEG and the electricity market operator. There are three types of certificates—for electricity savings, for gas savings, and for primary energy savings—and they are fully fungible.

In France, any economic actor can make savings actions and get certificates as long as the savings are at least 3 GWh over the lifetime of a measure. Certificates are delivered after the programs are carried out but before the realization of energy savings.[1]

In Great Britain, there are no certificates in the strict sense of the word. The scheme covers obliged parties, and no other party can receive verified savings that can be used to demonstrate compliance with the savings target. Suppliers may trade among themselves either energy savings from approved measures or obligations with written agreement from the regulator. There has been little interest in trading to date because energy savings can be traded only after the supplier's own energy-saving target has been achieved. Suppliers are also allowed to trade excess energy savings into the national emission trading scheme as carbon savings. However, the linking of carbon savings to the national emission trading scheme was never formalized. Suppliers have been allowed to carry savings over from EEC-1 to EEC-2, and this is what all suppliers that exceeded their target have chosen to do.

CONCLUSION

This entry discussed the general concept and the key elements of a system with energy-saving targets and tradable certificates for energy savings. It provided an up-to-date review of white-certificate schemes as implemented in three European countries, discussing some key design and operational features such as projects, implementer, and technology eligibility, and it pointed out key issues such as additionality, baseline setting, and M&V. This entry also explored the implications of different certificate trading rules and how these can affect the actual structure of the certificate market. Because the implementation track record of white-certificate schemes is very limited, it remains to be seen whether this policy portfolio will perform as expected, at what cost this will be achieved, and whether it can co-exist with and complement other MBIs in the energy sector to pave the road to a sustainable energy future.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contribution of numerous reviewers who have provided comments, suggestions, updated country information, and clarifications at various stages of the writing of the report "Tradable certificates for energy savings (white certificates): theory and practiceon which this entry is based."[3]

REFERENCES

1. Baudry, P.; Monjon, S. *The French Energy Law and White Certificates System*, RECS open seminar, Copenhagen, DK, 25 September, 2005.
2. Bertoldi, P.; Rezessy, S. Tradable certificates for energy efficiency: the dawn of a new trend in energy policy?

American Council for Energy Efficient Economy (ACEEE) Summer Study on Energy Efficiency in Buildings, ACEEE: Asilomar, CA, 2004.
3. Bertoldi, P.; Rezessy, S. Tradable certificates for energy efficiency (white certificates): theory and practice In *EUR 22196 EN*; Institute for Environment and Sustainability, DG Joint Research Centre, European Commission: Ispra, VA, 2006.
4. Bertoldi, P.; Rezessy, S.; Bürer, M.J. Will emission trading promote end-use energy efficiency and renewable energy projects? *Summer Study on Efficiency in Industry of the American Council for Energy Efficient Economy*, ACEEE: New York, 2005.
5. Bertoldi, P.; Rezessy, S.; Langniss, O.; Voogt, M. White, green and brown certificates: how to make the most of them? *Summer Study of the European Council for Energy Efficient Economy*, ECEEE: Mandelieu, 2005.
6. Jaccard, M.; Mao, Y. Making markets work better. In *Energy for Sustainable Development. A Policy Agenda*; Goldemberg, J., Ed.; UNDP: New York, 2002; 41–76.
7. Langniss, O.; Praetorius, B. How much market do market-based instruments create? An analysis for the case of "white" certificates. *European Council for Energy Efficient Economy (ECEEE) Summer Study*, ECEEE: Saint Raphael, France, Stockholm, 2003.
8. Lees, E. *How Increased Activity on Energy Efficiency Lowers the Cost of Energy Efficiency*; RECS seminar: Copenhagen, DK, 2005.
9. Lees, E. Summary of workshop. *Bottom-up Measurement and Verification of Energy Efficiency Improvements: National and Regional Examples, DG TREN, European Parliamnet*, ECEEE: Brussels, 2005.
10. Mansero, S. *Energy Efficiency Commitment. An Example From the U.K.*; RECS International: Copenhagen, DK, 2005.
11. Oikonomou, V.; Patel, M.; Mundaca, L.; Johansson, T.; Farinelli, U. A quallitatiive analysis of white, green certificates and EU CO2 allowances. *Phase II of the White and Green Project*, Utrecht University, Copernicus Institute: Utrecht, 2004.
12. Pavan, M. What's up in Italy? Market liberalization, tariff regulation and incentives to promote energy efficiency in end-use sectors. *American Council for Energy Efficient Economy (ACEEE) Summer Study*, ACEEE: Asilomar, CA, 2002.
13. Pavan, M. Expectation from Italy: suggestions to focus the project proposal. *Workshop on Organising a Task on White Certificates Under the IEA-DSM Implementing Agreement*, CESI: Milan, 2003.
14. Pavan, M. The Italian White Certificates System: Measurement and Verification Protocols. *EU and ECEEE expert seminar on Measurement and Verification in the European Commission's Proposal for a Directive on Energy Efficiency and Energy Services*, Brussels, 2004;Available at http://www.eceee.org/european_directives/EEES/monitoring_evaluation/ESD21SeptPavan.ppt, (accessed on December 2006).
15. Pavan, M. *Italian Energy Efficiency Obligation and White Certificates: Measurement and Evaluation*, Copenhagen, DK, 23 September, 2005.
16. Pavan, M. Italian Energy Efficiency Obligation and White Certificates: Measurement and Evaluation. *European Parliament Workshop on Case studies of Current European Schemes for the Measurement and Verification of Energy Efficiency Improvements*, Brussels, 2005;Available at http://www.danskenergi.net/attachments/8564/1146129527/marcellapavan.pdf, (accessed on December 2006).
17. Politechnico di Milano, *Final report of SAVE project, DSM pilot actions, DSM bidding and removal of DSM disincentives in restructured electricity markets*, 2000.

Transportation Systems: Hydrogen-Fueled☆

Robert E. Uhrig
University of Tennessee, Knoxville, Tennessee, U.S.A.

Abstract
The term "hydrogen economy" is the title of a recent book (Rifkin, J. *The Hydrogen Economy*, Tarcher/Putham Publisher, ISBN 1-58542-193-6, 2002.) but the concept of using hydrogen as fuel for transportation systems has been advocated by environmentalists and others for at least three decades. There is no universally accepted definition of "hydrogen economy," but it is generally viewed as the replacement of the vast majority of petroleum fuels used by transportation vehicles of all kinds (automobiles, trucks, trains, and aircraft) with hydrogen that is burned in internal-combustion engines, external-combustion (jet) engines, or preferably, used in fuel cells to more efficiently generate power for transportation. This chapter reviews the hydrogen economy from a basic energy engineering viewpoint to identify potential impediments and opportunities and to quantify the tasks involved in terms of energy and materials required in bringing the hydrogen economy into existence.

PRESENT TRANSPORTATION ENERGY SITUATION

Today, the United States uses over 20 million barrels (a "barrel" as used in the petroleum industry, is a volume of 42 gallons, which is equal to ~159 l) of oil per day (Mbbl/day), of which about 13 Mbbl/day are used for transportation of all kinds—cars, trucks, aircraft, trains, and military vehicles. If each barrel of oil contains 5.8 million British thermal units (MBtu) [6.12×10^9 Joules (J)], then the transportation energy to be replaced is about 75.4×10^{12} Btu/day [79.5×10^{15} J/day]. Because the energy content of hydrogen is about 51,600 Btu/lb [120×10^6 J/kg], the amount of hydrogen required to replace transportation fuel would be about 1.46×10^9 lb/day [664×10^6 kg/day]. If we assume that half of the transportation energy goes to fuel cells that are twice as efficient as heat engines while the remaining half is burned in combustion engines with a 20% increase in efficiency from current engines, then the required hydrogen is reduced to about 0.97×10^9 lb/day [441×10^6 kg/day], or about 177 million tons per year [~161 million tonnes (a 'tonne" is a metric ton, defined as 1000 kg or about 2200 pounds) per year]. This compares with the current production of about 50 million tons [~45.5 million tonnes] of hydrogen per year worldwide for all purposes, including fertilizers, upgrading of hydrocarbon fuels, and chemical-industry feedstock.

The Challenges of Hydrogen

Like electricity, hydrogen is an energy carrier, not an energy resource, and thus it must be extracted from other sources such as water or hydrocarbon fuels. Hydrogen is difficult to store, particularly on transportation vehicles, and difficult to distribute from one location to another. Hydrogen also burns with an almost invisible flame and is subject to special handling regulations by state and national codes for safety reasons. Clearly, the use of hydrogen as a transportation fuel has many engineering challenges that may be difficult to address. Despite this, there is a particular focus on hydrogen as an energy resource because it can be converted into electricity for transportation using fuel cells with an efficiency that is about twice as high as the conversion in conventional thermodynamic heat engines. Other benefits include the very significant reduction of pollutants and greenhouse gases being emitted into the atmosphere by transportation vehicles using hydrogen as a fuel.

HYDROGEN PRODUCTION

The two primary sources of hydrogen today are the decomposition of water by electrolysis and themochemical water-splitting processes and extraction from hydrocarbon fuels by chemical processing. Extracting hydrogen from water is more energy intensive than extracting it from hydrocarbon fuels. Coal gasification and partial oxidation of heavy oils are also demonstrated technologies for producing syngas (CO and H_2) from which hydrogen can be separated, but the large amount of CO_2 that is also

☆ The material in this article was originally published as *Engineering Challenges of the Hydrogen Economy* by Robert E. Uhrig in the Spring 2004 issue of *The Bent of Tau Beta Pi*, quarterly publication of the Engineering Society, Volume XCV, No. 2, Pages 10–19, 2004. That original article has been condensed and modified to reflect changes that have occurred since its publication.

Keywords: Hydrogen economy; Electrolysis; Steam methane reforming; Sulfur iodine process; Hydrogen storage; Hydrogen transportation and distribution; Hydrogen safety.

produced is a greenhouse gas and hence an undesirable by-product.

STEAM-METHANE REFORMING

Today, the vast majority of hydrogen is produced by steam methane reforming (SMR) of natural gas (which is about 95%–98% methane—CH_4) followed by a water–gas shift reaction. The two-step SMR process can be represented by

$$CH_4 + H_2O + \text{Heat} \xrightarrow{\text{Catalyst}} CO + 3H_2$$

Steam reforming reaction

$$CO + H_2O + \text{Heat} \xrightarrow{\text{Catalyst}} CO_2 + H_2$$

Water–gas shift reaction

The hydrogen comes from the methane and the steam. The SMR reaction that takes place at 300 psi [2.07 Mega-Pascals] and 1562°F [850°C] is endothermic with the required heat energy normally produced by combustion of some of the methane and the exothermic heat of the water–gas shift reaction. A well-designed SMR plant will yield hydrogen having about 75%–80% of the energy of the methane supplied. Unfortunately, SMR produces CO_2 in both the methane combustion and in the water–gas shift reaction. Recent work in Japan has demonstrated the feasibility of substituting high-temperature heat from a gas-cooled nuclear reactor to replace the heat supplied by the combustion of methane. This reduces the amount of methane required by about 20%–25% and eliminates the CO_2 produced by its combustion but not the CO_2 produced by the water–gas shift reaction.

The major disadvantages of steam methane reforming in the United States is the inadequacy of domestic supplies of natural gas, which produces wide fluctuations of price and the fact that it produces significant carbon dioxide that goes into the atmosphere. Reduction in the amount of CO_2 produced and the virtual elimination of pollutants are major benefits of using hydrogen as a transportation fuel in a "hydrogen economy." However, using a method such as steam methane reforming that emits CO_2 to produce hydrogen compromises the environmental benefits of using hydrogen as an automotive fuel. Perhaps more importantly, the demand for natural gas is growing rapidly primarily because of its increasing use as fuel for high-efficiency combined cycle gas turbine plants to generate electricity, as evidenced by the fact that since 2000 the average amount of natural gas used for electrical generation has exceeded the total amount used for home heating even though the vast majority of new homes use natural gas for heating. To produce hydrogen for the mature U.S. hydrogen economy of about 177 million tons per year [$\sim 161 \times 10^6$ tonnes/year], using SMR would require about 22.8×10^{12} standardized cubic feet [$\sim$707 billion normalized cubic meters] of natural gas per year in the United States.[10] This exceeds the average of about 21.0×10^{12} standardized cubic feet [651 billion normalized cubic meters] of natural gas per year currently used in the United States for all purposes (feedstock for chemical processes, upgrading hydrocarbon fuels, fertilizer, home heating, and generation of electricity) today. Gas utility executives are already planning importation of liquefied natural gas (LNG) to help meet the need in the United States. However, there are serious public concerns about the safety of LNG terminals. Doubling the U.S. demand for natural gas to meet the needs of the hydrogen economy is generally viewed as not being a viable alternative. This leads to the conclusion that SMR is not likely to be used to produce large quantities of hydrogen for the hydrogen economy.

ELECTROLYSIS OF WATER

Electrolysis of water is a mature technology used to produce high-purity hydrogen with an overall efficiency of about 25%. Commercial units of 10 MWe capacity are available today that can be coupled together to produce a facility of any desired size. On average, 1 MW of electricity using today's technology will produce about 1040 pounds (lb) [473 kilograms (kg)] of hydrogen per day and 8320 lb [3747 kg] of oxygen per day from about 10,000 lb [$\sim$4550 kg] of water per day (including $\sim$6% overall processing loss of water).[9] Hence, to produce 177 million tons [$\sim$161 million tonnes] of hydrogen per year for a mature hydrogen economy would require about 932 Gigawatts of electricity (GWe), i.e., 927 new nonpolluting 1000 MWe power plants, more than the total generating capacity of the United States today, and about 1700 million tons [$\sim$1545 million tonnes] of water per day.

Use of "Spinning Reserve" to Produce Hydrogen

One of the useful features of conventional electrolysis (producing hydrogen using electricity only) is the ability to instantaneously disconnect the electrolyzer from the utility electric service and use this available electricity for other purposes. This allows the amount of capacity reserved by the utility to instantaneously pick up the electrical load dropped because of plant or transmission system failure (commonly called "spinning reserve") to be used for electrolysis. In effect, this increases the useful generating capacity of the utility by the amount of spinning reserve used for electrolysis. The additional cost to the utility for this additional capacity is only the cost of fuel and maintenance because the capital charges are already covered. For instance, a utility with a total generating

Ther–Und

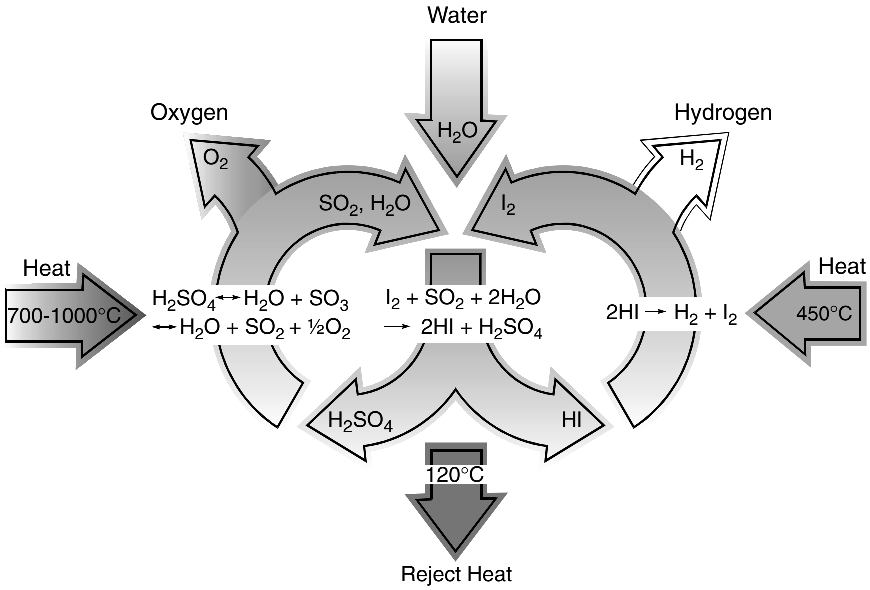


Fig. 1 Schematic diagram of the sulfur iodine process of producing hydrogen courtesy of oak ridge national laboratory. Source: From OECD Nuclear Energy Agency (see Ref. 3).

capacity of 12,000 MWe with its largest plant having a 1200 MWe output would have to carry 1200 MWe of "spinning reserve" to pick up the customer load in the event that the 1200 MWe plant trips offline. Hence, the utility's useful capacity is only 10,800 MWe. If this utility were able to use this spinning reserve to generate hydrogen by electrolysis while still being able to use it for an emergency, the only additional cost would be operating costs (fuel and maintenance). Although cost accountants might want to assign the capital costs equally to all activities, the rationale for not assigning capital costs to the electrolysis is that this capacity cannot be used for any other activity and hence does not increase the cost of electricity produced for other uses.

A study of the price of electricity on the Alberta, Canada Open Market by Atomic Energy of Canada Limited showed that there was an underlying low cost with many cost spikes. These spikes were 10–20 times as high but occurring for only about 5% of the time. Hence, for about 95% of the time, the average cost of electricity is below the overall average cost.[5] The ability to instantaneously switch off the hydrogen electrolyzers during peak loads could significantly reduce the cost of producing hydrogen by electrolysis. Similar pricing patterns exist in the United States and may exist in other countries.

High-Temperature Electrolysis

Electrolysis is more energy intensive than other methods of producing hydrogen. The energy of the hydrogen produced by electrolysis is about 75% of the energy of the electricity used, which in turn, typically has about one-third of the energy of the fossil or nuclear fuel used to generate the electricity. Hence, the overall efficiency of electrolysis is about 25%. There are two ways to improve this overall efficiency: (1) use more efficient high-temperature power plants to generate the electricity and (2) use high-temperature electrolysis.

The use of very high-temperature heat (800°C–900°C) with electricity for electrolysis is being investigated for electrolysis of steam. Substitution of high-temperature heat energy for some of the electrical energy in electrolysis has been demonstrated to decrease the electricity needed and appears to be potentially competitive with other hydrogen production processes under some circumstances. About 30% of the energy is heat that is less expensive than electrical energy. The overall efficiency of high-temperature electrolysis is 45%–50%, higher than conventional electrolysis.[3] Using high-temperature cycle electrical generation with high-temperature electrolysis could further improve the efficiency.

THERMOCHEMICAL CRACKING OF WATER

"Water-Splitting"—The Sulfur–Iodine Process

Thermochemical cracking of water is generally viewed as the primary way that nuclear energy will be used to produce hydrogen on a large scale. Thermochemical cycles are Carnot cycle-limited, which indicates that high temperatures can improve the conversion efficiency. A comprehensive study of potential thermochemical cycles for hydrogen production reactions by the International Atomic Energy Agency (IAEA) included metal oxide systems, metal-halide family processes, sulfur family processes, and numerous hybrid systems incorporating portions of two or more of the basic systems.[4] Only a few of these cycles have survived the comprehensive

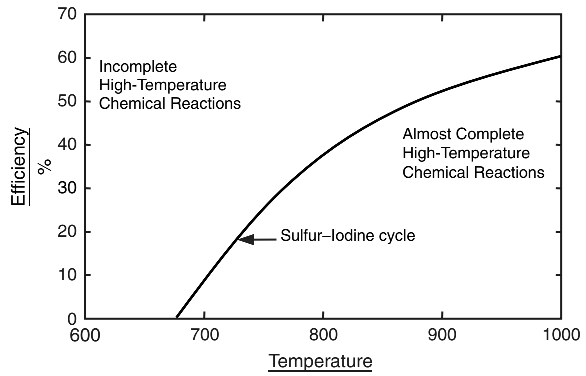


Fig. 2 Efficiency of Sulfur–Iodine cycle as a function of temperature (°C). Courtesy of General Atomics.
Source: From Ref. 8.

reviews and studies of the past two decades. When all of the steps involved in any given thermochemical process are summed up, the result is a closed system represented by

$$H_2O + \text{high-temperature heat} \rightarrow H_2 + \frac{1}{2}O_2 + \text{low-temperature heat}$$

In theory, this reaction could be implemented directly, but it would require a very high temperature (at least in the 2500°C–3000°C range), which is well beyond the capability of any known nuclear reactor.

General Atomics (GA) has studied the potential use of helium-cooled nuclear reactors to produce hydrogen for over three decades. After reviewing 922 references and 115 cycles, some 25 cycles were investigated in detail using thermodynamic calculations and preliminary block flow diagrams. Only two were selected for detailed design studies: the sulfur–iodine (S–I) cycle and the Adiabatic UT-3 cycle. General Atomics chose the S–I process for further detailed study and analysis. The fact that the S–I cycle involved only fluids (liquids and gases), the higher efficiency of the S–I cycle, and the compatibility between the energy requirements of the S–I cycle and the characteristics of its helium-cooled reactors were some of the reasons for this choice.[8]

Sulfur–Iodine (S–I) Cycle

Probably the most studied cycle is the sulfur–iodine cycle, one of three sulfur cycles being considered today. All three sulfur cycles have a common thermal decomposition of sulfuric acid that requires a very high temperature (~900°C, 1652°F) and a catalyst. The basic sulfur–iodine process that has been studied extensively by GA since the 1970s[7] and more recently at Oak Ridge National Laboratory and in Japan is shown schematically in Fig. 1.[3] The high-temperature reaction shown on the left side of Fig. 1 is represented by

$$H_2SO_4 \overset{(600°C)}{\leftrightarrow} SO_3 + H_2O \overset{(850°C \text{ and a catalyst})}{\leftrightarrow} SO_2 + H_2O + \frac{1}{2}O_2$$

The double-headed arrows indicate that the reactions can go either way depending upon temperature, pressure, and material balance. This high-temperature reaction should go almost to completion for efficient hydrogen production. This is shown graphically in Fig. 2, where the efficiency approaches 60% when the temperature is 1000°C (1832°F).[8]

The H-I process has two lower temperature reactions—(1) the hydrogen-generating reaction shown on the right side of Fig. 1 (where only about 16% of the HI is decomposed with the remainder being recycled), represented by[2]

$$2HI \overset{(350°C\text{–}450°C)}{\leftrightarrow} I_2 + H_2$$

and (2) the Bunsen reaction shown in the middle of Fig. 1 that involves recycle of the chemicals and the addition of water, represented by

$$I_2 + SO_2 + 2H_2O \overset{(120°C)}{\leftrightarrow} 2HI + H_2SO_4$$

This reaction requires excess iodine so that the hydrogen iodine and sulfuric acid separate into two separate liquids to be recycled in the process. The management of the heat recovery throughout these processes is critically important to the high efficiency of all of the sulfur cycles.

The advantages of the S–I cycle are (1) all of the chemistry reactions have been demonstrated, at least on a laboratory scale, (2) only fluids (liquids and gases) are involved in continuous processes, and (3) all chemicals are recycled. While there have been extensive studies of the various cycles in the United States, the only experimental testing of the whole SI cycle has been in Japan, where successful production of hydrogen has been demonstrated on a bench scale using a nonnuclear high-temperature heat source.

Two other sulfur cycles that are being studied are: (1) the “Hybrid Sulfur” Cycle (known as the GA 22, Westinghouse, and Ispra Mark 11 cycles) and (2) the “Ispra Mark 13 Cycle.” Like the S–I cycle, both of these cycles involve the thermal decomposition of sulfuric acid that requires a high temperature (~900°C) and a catalyst but does not involve iodine. In the Hybrid Sulfur cycle, the two low-temperature processes of the S–I cycle are replaced by electrolysis that takes place at 90°C (194°F), and can be represented by

$$SO_2 + 2H_2O \overset{\text{Electrolysis}}{\leftrightarrow} H_2SO_4 + H_2$$

Electrolysis is also used in the Ispra Mark 13 cycle in conjunction with a second chemical process involving bromide to enhance the electrolysis.

Reducing the Temperature Required for Cracking Water

Detailed studies have concluded that peak temperatures in the S–I cycle need to be at least 850°C (1562°F) and preferably 900°C (1652°F) to drive the SO_3 decomposition to near completion.[3] Hence, the reactor outlet temperature would need to be about 950°C (1742°F) to 1000°C (1832°F). The High-Temperature Test Reactor (HTTR) helium-cooled reactor in Japan has a 950°C (1742°F) outlet temperature, the same as the German AVR pebble bed test reactor in the 1970s.

Studies of the process flow sheets show that there are strong economic and other incentives to lower the temperature and increase the pressure at which SO_3 dissociates—the exact opposite of the conditions required by thermodynamic considerations. After the high-temperature dissociation reaction, all the chemicals must be cooled to near room temperature, the SO_2 must be separated out and sent to the next chemical reaction, and the unreacted H_2SO_4 (formed by recombination of SO_3 and H_2O at lower temperatures) must be reheated back to high temperatures. Unless the chemical reactions go almost to completion, the energy losses in separations and in heat exchangers to heat and cool all the unreacted regents (H_2SO_4) could result in an inefficient and uneconomical process.[2]

Recent work at Oak Ridge National Laboratory (ORNL) has been directed at accelerating the SO_3 disassociation into SO_2 and O_2 at lower temperatures and higher pressures by the use of an inorganic separation membrane.[3] It appears to be possible to lower the required temperature to about 700°C (1292°F) by separating and removing the SO_2, H_2O, and O_2 products from the SO_3 disassociation. If the reaction products are removed, the remaining SO_3 (with a catalyst and heat) will disassociate more rapidly. If the reaction products can continue to be selectively removed, the chemical reaction can be driven to completion at a lower temperature, estimated to be ~700°C (~1292°F). The membrane operates with high pressure on one side and lower pressure on the other and this pressure difference dives the separation process. Preliminary results of tests at ORNL have been encouraging.[3]

Lowering the high-temperature requirements for the SI processes has many advantages. Foremost of these advantages is that currently designed high-temperature reactors, i.e., some existing high-temperature gas-cooled reactors as well as liquid metal (lithium or lead) cooled fast reactors, could be used for the nuclear production of hydrogen. Furthermore, lower temperatures significantly enhance the safety of both the reactor and the hydrogen production facilities by reducing peak temperatures in structural materials to levels where there is a better understanding of their mechanical properties, particularly strength and creep behaviors.

Calcium Bromide Cycles (UT-3 and Others)

A family of lower temperature calcium–bromide cycles for producing hydrogen with a 750°C (1382°F) peak temperature is also being studied. The high-temperature steps are well understood, but they involve solid reagents, calcium oxide (CaO), and calcium bromide ($CaBr_2$). The dimensions of the solids change size when the chemical reactions occur and there is the potential for dust. Several options are being studied as methods of producing hydrogen and recycling the bromide, including (1) the Iron Cycle process (UT-3) that was developed at the University of Tokyo in Japan, (2) electrolysis that has a low efficiency, and (3) a cold plasma process that provides energy as a replacement for electricity in electrolysis.[3]

Status of the Development of Thermochemical Cycles in the United States.

The United States Department of Energy's Nuclear Hydrogen R&D plan has as its goal to define the hydrogen production R&D necessary to demonstrate nuclear hydrogen on a production scale by 2017. Currently, they are supporting studies of several of the thermochemical cycles discussed in this chapter as well as experimental work on some of them. The S–I cycle is currently being developed experimentally by three organizations: Sandia National Laboratory (SNL), GA, and the French Commissariat a l'Energie Atomique (CEA). Sandia National Laboratory is building the facilities to study and test the high-temperature thermal decomposition of sulfuric acid using a nonnuclear source of high-temperature heat, GA is investigating the hydrogen generating reaction, and CEA is investigating the Bunsen reaction that involves recycle of the chemicals and the addition of water.[8] When this work is completed, a prototype system integrating these three processes in a single facility will be undertaken. Current DOE plans are to demonstrate both the S–I thermochemical water-splitting and high-temperature electrolysis processes on a production scale at Idaho National Laboratory (INL) by 2017. The INL facility is expected to provide 60 MWt of high-temperature heat for hydrogen production and to demonstrate passive safety while maintaining the purity of the hydrogen to meet fuel cell standards.

HYDROGEN STORAGE

Implementation of the hydrogen economy will require facilities for storing, transporting, and distributing hydrogen to refueling facilities throughout the country.

Historically, hydrogen has been stored as a high-pressure gas (~3000 psi) or a cryogenic liquid (~20°K). Although there are experimentations with liquid hydrogen by some automakers, most current discussions about storage of hydrogen on automotive vehicles involve gaseous hydrogen at 5000 psi or possibly 10,000 psi. The liquefaction of hydrogen consumes about 30% of the energy stored, and there is also a continual loss of energy due to thermal conduction through the insulated walls whether the liquid is stored in a service station or a vehicle on or off the road. Even so, there may be applications for liquid hydrogen in heavy vehicles and long-range aircrafts, where weight is critical.

Storage of Hydrogen as a High-Pressure Gas

An official goal for on-board hydrogen storage is to achieve a 300-mi range with a tank no larger than current automobile fuel tanks. The importance of the goal is reflected in the general belief that failure to meet this goal was a major impediment to battery-powered electric cars—a technology in which the direct storage of electricity in batteries is much simpler than converting electricity to hydrogen, distributing it nationally, dispensing it to vehicles, and then converting hydrogen to electricity in a fuel cell to drive an electric motor to propel a vehicle.

On-board storage tanks for gaseous hydrogen on vehicles, made with filament-wound carbon fibers and lined with aluminized polyester bladders, have been approved for use up to 5000 psi (34.5 MP) by United States and up to 10,000 psi (69 MP) by German authorities. Generally, these tanks are cylindrical, and more than one are sometimes used because of the required volume and the low-volumetric-energy density of hydrogen. Even though the higher efficiency of fuel cells partially compensates for this low-energy density, the large size of the tank(s) required for a 300-mi range is a concern and researchers still seek alternatives. Technologies under serious consideration are metallic hydrides and alanates where hydrogen is adsorbed onto interstitial surfaces. Reports on these two options from a recent conference are summarized below. Adsorption of hydrogen by carbon nanotubes was considered as an option a few years ago, but it has not lived up to its initial promise for hydrogen storage.

Metal Hydrides for Storage of Hydrogen

Hydrogen is a highly reactive element and it will form hydrides and solid solutions with hundreds of metals and alloys as well as form chemical compounds or complexes with many other elements. Metal hydrides are formed by hydrogen bonding to a metal with metallic, ionic, or covalent bonds. Hydrogen is usually bound in the interstitial sites, and it can be removed by applying heat to the metal hydride. Many intermetallic compounds and solid solutions can readily absorb and desorb hydrogen gas at room temperature and atmospheric pressure, but these hydrides can reversibly store only 1–3 weight-percent hydrogen, which is not enough for application to vehicle storage. Covalent and ionic hydrides (e.g., MgH_2 and LiH, respectively) are capable of storing 7–12 weight-percent hydrogen, but they must be heated to above ~325°C (617°F) to release the hydrogen at atmospheric pressure. This temperature is much higher than the ~75°C (167°F) waste heat that is available from a proton-exchange-membrane fuel cell. Recent work with catalyzed complex hydrides containing mixed ionic-covalent bonding can reversibly store more than 4% hydrogen with operating temperatures below 125°C (257°F). Generally, the volumetric energy density of hydrogen stored in metal hydrides is comparable to that of liquid hydrogen. The primary problems with metal hydrides for vehicle applications are their heavy weight and high cost.[1]

Alanates for Hydrogen Storage

Alanates, such as sodium alanate ($NaAlH_4$), are aluminum alloys that contain hydrogen. Sodium alanate undergoes two-step decomposition into sodium hydride (NaH), hydrogen, and aluminum, where the gas temperature and pressure determine the equilibrium quantities of reactants and products that release a total of about 5.6% hydrogen. Pure alanates react slowly at ambient conditions, but recent work indicates significant improvement in sorption kinetics by adding transition metal additives. Many other alanates show promise as a storage medium, but they are at an early stage of their development.[1]

HYDROGEN DISTRIBUTION

Because of its low volumetric energy density, hydrogen is at a disadvantage when compared to other gaseous fuels with higher volumetric energy densities (e.g., methane or propane) and liquid fuels (gasoline or ethanol). The power to perform the pumping of hydrogen is reported to be about a factor of 4.6 greater than for methane over a range of high pressures. The total cost of distribution for an equal amount of energy in the form of hydrogen is estimated by the International Energy Agency to be 15 times that of liquid hydrocarbons. However, the economies of scale associated with large central hydrogen facilities may partially compensate for the extra distribution costs.

There are perhaps two basic but different configurations of a distribution system that may evolve over the next three decades—with many combinations of the two and other variants. Only the two basic configurations will be discussed here.

Hundreds of Large Hydrogen Plants with a National Distribution Pipeline Grid

In this scenario, there are many clusters of large hydrogen plants distributed around the country with an interconnected pipeline grid to distribute the hydrogen to control centers, which in turn carry hydrogen to millions of service stations through a distribution pipe grid. Most likely, the plants would be thermochemical water-splitting plants with nuclear power plants providing the heat. This arrangement is directly analogous to the current electric transmission and distribution grids in the United States.

If we accept the required hydrogen production capacity as about 177 million tons (161 million tonnes) per year for the hydrogen economy (this is based on present-day U.S. transportation sector oil demand), which is equivalent to about 145 billion normalized cubic feet (Mncf/day) [4.50 billion standard cubic meters per day (Mscm/day)], and that today's world-class hydrogen plants each produce about 184 Mncf/day [5.7 Mscm/day], then 790 world-class hydrogen production plants would be required for the hydrogen economy. These plants might be grouped in 197 clusters of four plants per cluster or an average of about four clusters per state. Clearly, more clusters would be needed in the more populous states and fewer plants would be needed in the other states. It can be readily shown that the 184 Mncf/day (5.7 Mscm/day) output of hydrogen is equivalent to ~775 MWt. Hence, if each world-class plant is operating at 50% efficiency (typical of thermochemical water-splitting plants), we would require ~1550 MWt per plant or 6200 MWt for each four-unit cluster and a total of ~1225 GWt of heat energy for 790 plants. A crude calculation indicates that if the 197 four-unit clusters of plants were laid out uniformly on a square grid throughout the United States, the clusters would be about 80 mi apart and would require about 90,000 mi of interconnecting pipe. These interconnecting pipes would be analogous to the high-voltage transmission lines for electricity. If a hydrogen distribution grid were installed with connections at 10-mi intervals on the interconnecting pipes, an additional 725,000 mi of smaller distribution pipe would be required.

Millions of Small Hydrogen Plants Located at Hydrogen Service Stations

The other basic arrangement for distribution of hydrogen to service stations is a distributed array of small hydrogen generators, probably electrolysis units, of sizes selected to provide adequate hydrogen for a local area's need and using electricity from the most economical source. In the hydrogen economy, the total amount of new electrical capacity required for electrolysis would be about 930 GWe—more than doubling the current electrical generating capacity of the United States. The electrical grid would also have to be doubled in size to carry the needed electricity.

The alternative configuration for providing the required power is a distributed array of windmills or photovoltaic generators. Some 930,000 new 1 MWe windmills would be required if they ran 24 h a day. With an average availability factor of about 30%, the number of windmills would have to be increased significantly. The other option, some 93,000,000 new 10 kWe photovoltaic electric generating units, is also burdened with a low-availability factor—only 20% and significant additional capacity would be required. Clearly, the ultimate source of power to provide hydrogen for the hydrogen economy would be some combination of all the options discussed above. There may be others, including fusion plants and fermentation-or solar-driven photo-biological methods that are currently being studied.

Hydrogen at Service Stations

There are several issues of concern associated with dispensing hydrogen into automotive vehicles. The first is that the connection from the station to the vehicle would be complex (with sensors to assure secure attachment without leakage). The valves and other hardware used to control the flow of very high-pressure gases are inevitably complex and expensive. To deliver hydrogen to vehicle tanks at 5000 psi will require a continuous delivery pressure of perhaps 7000 psi to keep the refilling time reasonable. As the station uses its supply of hydrogen, the pressure will drop. To keep the pressure high enough to refuel vehicles in a reasonable time would require a compressor to deliver hydrogen into the vehicle.

EPILOGUE

The National Environmental Policy Act, passed by Congress in 1969, mandated that before a major project can be undertaken, an environmental impact statement (EIS)—a comprehensive assessment of the benefits of the project and its impact upon all aspects of the environment—must be prepared and presented to the authorizing authority. Nothing remotely resembling an environmental impact statement required for the construction of a nuclear power plant or other large projects (an EIS for off-shore oil drilling had ~32,000 pages) has been undertaken for the hydrogen economy. It would seem that, given the extraordinary magnitude of the requirements for power, water, fuels, and infrastructure of all sorts, that at least a scoping study of the environmental and economic consequences of the hydrogen economy should be undertaken now.

REFERENCES

1. Bowman, R. Jr.; Sanrock, G. *Hydrides for Energy Storage and Conversion Application*, Proceedings of GLOBAL 2000 Atoms for Prosperity, American Nuclear Society Meeting, New Orleans, LA, November 16–20, 2002.
2. Forsberg, C. *Production of Hydrogen Using Nuclear Energy*, Presentation to National Research Council/ National Academy of Science: Washington, D.C., January 22, 2003.
3. Forsberg, C.; Bischoff, B.; Mansur, L.; Lee Trowbridge, L. *Nuclear Thermo-Chemical Production of Hydrogen with a Lower-Temperature Iodine-Westinghouse-Ispra Sulfur Process*, Second Information Exchange Meeting on Nuclear Production of Hydrogen, OECD Nuclear Energy Agency: Argonne, IL, October 2–3, 2003.
4. IAEA. International Atomic Energy Agency document. Hydrogen as an Energy Carrier and Its Production by Nuclear Power, IAEA-TECDOC-1085, May, 1999.
5. Miller, A.; Duffy, R. *Hydrogen and High Temperature Reactors: a Source Map*, Proceedings of GLOBAL 2000 Meeting, at the Winter American Nuclear Society Meeting, New Orleans, LA, November 16–20, 2003.
6. Rifkin, J. *The Hydrogen Economy*, Tarcher/Putnam Publisher, ISBN 1-58542-193-6, 2002.
7. Schultz, K. *Efficient Production of Hydrogen from Nuclear Energy*, Presentation to the California Hydrogen Business Council, June 27, 2002.
8. Schultz, K. *Fueling the Hydrogen Economy: Hydrogen Production using Nuclear Energy*, Presentation at World Nuclear Association Meeting, October, 2003.
9. Stuart, S. *Nuclear-Electrolysis Synergies: For Neat Hydrogen Production and Cleaner Fossil Reserves Harvesting*. ANS President's Special Session: Hydrogen Systems, ANS Meeting: Reno, NV, November 11–15, 2001.
10. Uhrig, R. *Engineering Challenges of the Hydrogen Economy*. The Bent of Tau Beta Pi, Vol. XCM, No. 2, Spring, 2004.

Transportation: Location Efficiency

Todd A. Litman
Victoria Transport Policy Institute, Victoria, British Columbia, Canada

Abstract
This chapter describes the concept of location efficiency (also called smart growth), which refers to land-use development patterns that maximize the ease with which people can obtain desired goods, services, and activities, and that minimize the need for physical travel. Location efficiency includes factors such as land-use density, land-use mix, roadway connectivity, and transportation system diversity. Improving location efficiency helps reduce transportation costs, including road and parking facility costs, consumer costs, traffic risk, energy consumption, pollutant emissions, and other environmental impacts. It also benefits people who are transportation disadvantaged and increases housing affordability.

INTRODUCTION

Land use and transportation are two sides of the same coin. Transportation affects land use, and land use affects transportation. Decisions that affect one also affect the other. As a result, it is important to coordinate transportation and land-use planning decisions so they are complementary. Land-use planning can help create more efficient transportation systems that reduce per-capita vehicle travel and energy consumption, for example. This is referred to as *location-efficient* development or smart growth.

To understand how land use affects travel patterns it is useful to consider the concept of accessibility. Accessibility (or just access) is the ability to reach desired goods, services, activities, and destinations—together called opportunities. A stepladder provides access to the top shelf in your kitchen. A store provides access to goods. Libraries, telephones, and the Internet provide access to information. Paths and roads provide access from one destination to another by walking, cycling, automobile, and bus.

Access is the ultimate goal of most transportation, excepting the small portion of travel in which movement is an end in itself (e.g., cruising, historic train rides, horseback riding, and jogging). Even recreational travel usually has a destination, such as a resort or a campsite. (Mobility as an end in itself is discussed later in this chapter).

Four general factors affect accessibility:

1. Mobility—that is, physical movement. Mobility can be provided by walking, cycling, public transit, ride sharing, taxi, automobiles, trucks, and other modes.
2. Mobility substitutes, such as telecommunications and delivery services. These services can provide access to some types of goods and activities, particularly those involving information.
3. Transportation system connectivity, which refers to the directness of links and the density of connections in a path or road network.
4. Land-use patterns—that is, the geographic distribution of activities and destinations. When real estate experts say "location, location, location," they mean "accessibility, accessibility, accessibility."

Keywords: Land use; Smart growth; Location; Energy.

LAND-USE IMPACTS ON TRANSPORTATION

Land-use factors (also called spatial development, community design, urban design, or the built environment) can affect transport activity in several ways.[1] When worksites are dispersed and located in areas without good walking and cycling facilities, for example, most employees will drive, but if the same businesses locate in commercial centers with good walking and cycling facilities, and with good transit services, a significant portion of employees will use alternative modes.

Planners increasingly realize the importance of integrating transportation and land-use decisions to increase accessibility and improve travel options, thereby reducing the amount of motor vehicle travel required to meet people's needs and serve economic activities. This can help achieve a variety of planning objectives, including reduced congestion, energy conservation, pollution and emission reductions, infrastructure cost savings, increasing household affordability, and improving economic opportunities for disadvantaged populations.

Specific land-use factors that affect transportation are described in the following sections.[2]

Density

Density refers to the number of people or jobs in a given area. Increased density tends to reduce per-capita automobile ownership and use, and to increase use of alternative modes. Fig. 1 shows how per-capita vehicle mileage tends to decline with density in U.S. urban areas. Many other studies find similar results.

Increased density tends to reduce per capita vehicle travel.

Density at both origins and destinations affect travel behavior. One study found that increasing urban residential population density to 40 people per acre increased transit use from about 2 to 7%, while increasing densities in commercial centers to 100 employees per acre resulted in an additional 4% increase in transit use, to an 11% total mode share.[3] Both work trips and shopping trips are affected by population and employment densities.

Land-Use Mix

Mixed land use (such as locating appropriate businesses and public services in or adjacent to residential areas) can reduce per-capita vehicle travel. It tends to reduce the distances that residents must travel for some services and allows more use of walking and cycling for such trips. It also can reduce some employees' commute distances (some residents may obtain jobs at nearby businesses), and employees who work in a mixed-use commercial area are significantly more likely to commute by alternative modes.

Roadway Design

Roadway design can affect travel behavior in several ways. A connected road network provides better accessibility than a conventional hierarchical road network with a large portion of dead-end streets. Increased connectivity can reduce vehicle travel by reducing travel distances between destinations and by improving walking and cycling conditions, particularly where paths provide shortcuts, so walking and cycling are relatively faster than driving. This also supports transit use.

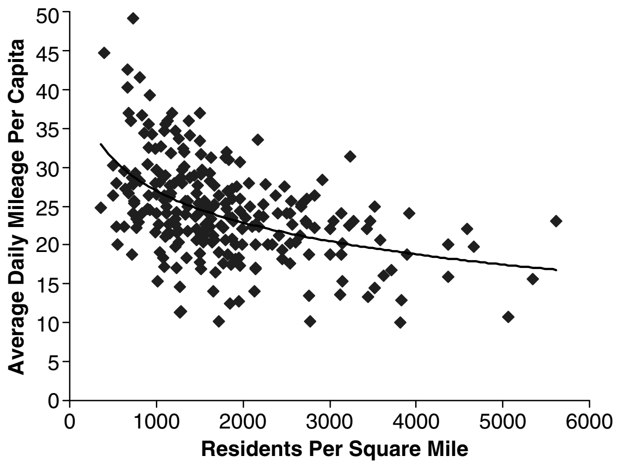


Fig. 1 Density vs vehicle travel for U.S. urban areas.
Source: From FHWA, 2005 (see Ref. 4).

Transit Service Quality

Per-capita automobile ownership and motor vehicle mileage tend to decline as transit service quality in an area improves. Transit service quality includes the convenience, frequency, comfort, and security of transit vehicles and stations or stops, as well as the quality of walking conditions in nearby areas.

Site Design and Building Orientation

Some research indicates that people walk more and drive less in areas with traditional pedestrian-oriented commercial districts where building entrances connect directly to the sidewalk than in areas with automobile-oriented commercial strips where buildings are set back and separated by large parking lots (Moudon, 1996).[5] This type of building orientation improves pedestrian access and creates a more attractive pedestrian environment.

Cumulative Impacts

The transportation effects of density and clustering, land-use mix, transit access, street design, and building design tend to be cumulative. As an area becomes more urbanized (more dense and mixed activities, higher land prices, and less parking), transportation diversity tends to increase, with fewer trips by automobile and a greater portion of trips by walking, cycling, and public transit.

Holtzclaw[6] finds that average vehicle ownership, vehicle travel, and vehicle expenditure per household decline with increasing residential densities and proximity to public transit, holding constant other demographic factors such as household size and income. The formulas below summarize his findings. An online calculator, This View of Density Calculator (www.sflcv.org/density), uses this model to predict the effects of different land-use patterns on travel behavior (Fig. 2).

This figure illustrates how density and transit accessibility affect household vehicle mileage. The Transit Accessibility Index (TAI) indicates daily transit service nearby.

Ewing, Pendall, and Chen[7] developed a sprawl index based on 22 specific variables related to land-use density, mix, street connectivity, and commercial clustering. The results indicate a high correlation between these factors and travel behavior: a higher sprawl index is associated with higher per-capita vehicle ownership and use, and lower use of alternative modes. Other studies also find

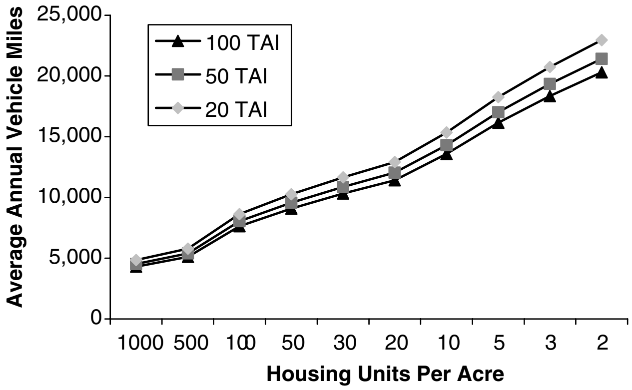


Fig. 2 Annual VMT per household.
Source: From Based on National Resources Defense Council (see Ref. 6).

that per-capita vehicle travel is significantly lower in higher-density, multimodal urban neighborhoods than in automobile-oriented suburban neighborhoods.

Lawton[8] found that average daily motor vehicle miles per adult decreased from 19.8 in the least urbanized residential neighborhoods to 6.3 in the most urban neighborhoods, due to fewer and shorter automobile trips. Even modest land-use changes can provide significant vehicle travel reductions if they are reinforced by other mobility strategies, such as commute trip reduction programs (which encourage commuters to use alternative modes) and parking pricing or cash-out (travelers can choose to receive cash instead of parking subsidies).

As a result, residents of more urbanized areas drive significantly fewer miles and rely more on alternative modes than residents of suburban and rural communities, as indicated in Fig. 3.

Urban residents drive less and use transit, cycling, and walking more than elsewhere.

LOCATION EFFICIENCY

Location-efficient development (also called smart growth) refers to more compact and mixed development located in compact centers designed for walking, cycling, and transit. Table 1 compares the two types of land-use patterns.

Per-capita motor vehicle travel tends to be significantly higher in automobile-dependent areas compared with the same demographic and businesses activity in Smart Growth locations.[10] Table 2 summarizes typical reductions in Vehicle Miles of Travel (VMT) resulting from various Smart Growth developments. This also indicates that location efficiency significantly reduces per-capita vehicle travel. Because of these impacts, planners are increasingly including land-use management strategies to achieve transportation planning objectives, including improved accessibility, reduced infrastructure costs, reduced accidents, and energy conservation.

HOW LOCATION-EFFICIENT DEVELOPMENT IS IMPLEMENTED

Location-efficient development is implemented by developers, usually with support and encouragement from local governments. It often is implemented as part of Smart Growth and New Urbanist planning.[11] In practice, location-efficient development consists of redeveloping older urban residential neighborhoods and commercial centers, creating new transit-oriented suburban neighborhoods, and improving walking conditions, cycling conditions, and transit services. It also can involve the application of mobility management (also called transportation demand management), which includes various policies and programs that encourage people to reduce their automobile travel and use alternative travel options.[13,10] These programs include incentives for more compact, infill development, and parking management to reduce the amount of parking required at each destination.

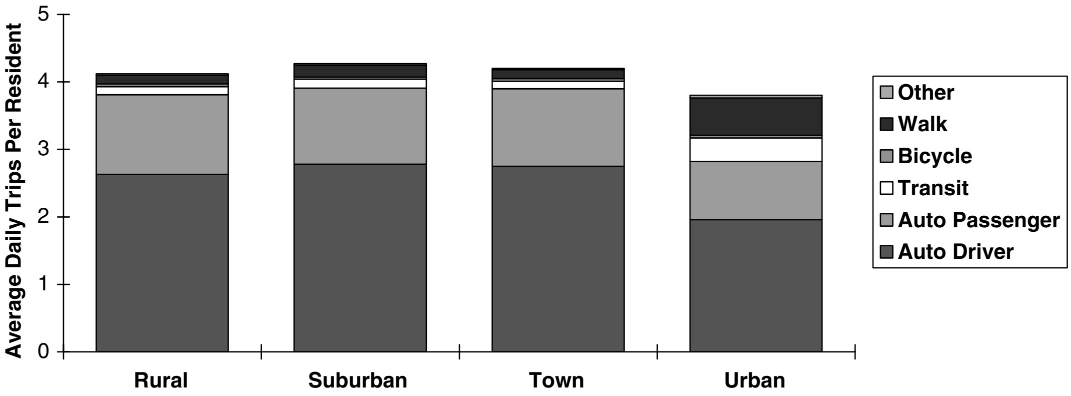


Fig. 3 Average daily trips per resident by geographic area.
Source: From NPTS (see Ref. 9).

Table 1 Comparing automobile dependency and smart growth

	Automobile dependency	Smart growth
Density	Lower-density, dispersed activities	Higher-density, clustered activities
Growth pattern	Urban periphery (greenfield) development	Infill (brownfield) development
Land-use mix	Homogeneous (single-use, segregated) land uses	Mixed land use
Scale	Large scale. Larger buildings and blocks, wide roads. Less detail, because people experience the landscape at a distance, as motorists	Human scale. Smaller buildings, blocks, and roads. Careful detail, because people experience the landscape up close, as pedestrians
Public services (shops, schools, parks)	Regional, consolidated, larger. Requires automobile access	Local, distributed, smaller. Accommodates walking access
Transport	Automobile-oriented transportation and land-use patterns, poorly suited for walking, cycling, and transit	Multimodal transportation and land-use patterns that support walking, cycling, and public transit
Connectivity	Hierarchical road network with numerous loops and dead-end streets, and unconnected sidewalks and paths, with many barriers to nonmotorized travel	Highly connected roads, sidewalks, and paths, allowing relatively direct travel by motorized and nonmotorized modes
Street design	Streets designed to maximize motor vehicle traffic volume and speed	Streets designed to accommodate a variety of activities. Traffic calming
Planning process	Unplanned, with little coordination between jurisdictions and stakeholders	Planned and coordinated between jurisdictions and stakeholders
Public space	Emphasis on the private realm (yards, shopping malls, gated communities, private clubs)	Emphasis on the public realm (streetscapes, pedestrian environment, public parks, public facilities)

Source: From Victoria Transport Policy Institute (see Ref. 12).

One strategy for encouraging households to choose more accessible locations is to offer location-efficient mortgages (LEMs), which means that lenders recognize the potential savings of a more accessible housing location when assessing a household's borrowing ability. It considers transportation and housing costs together, so vehicle cost savings are treated as additional income that can be spent on a mortgage. This gives home buyers an added incentive to choose location-efficient residences and tends to encourage more infill development as opposed to more automobile-dependent development at the urban periphery.[15–17]

Location-efficient mortgages are implemented by residential mortgage lenders, often with the support and encouragement of government agencies such as Fannie Mae and the Canadian Mortgage and Housing Corporation. Lenders use a model to determine which locations have lower transportation costs and, therefore, can qualify for higher mortgage payments. The following factors can be considered in such models:

- Residential density
- Land-use mix—that is, the number of public services within convenient walking distance (schools, shops, parks, medical services, pharmacy, etc.).
- Proximity to high-quality transit (such as a rail transit station or a bus line with frequent service).
- Quality of walking and cycling conditions.

Table 2 Infill VMT reductions

Location	Description	VMT reduction (%)
Atlanta	138-acre brownfield, mixed-use project	15–52
Baltimore	400 housing units and 800 jobs on waterfront infill project	55
Dallas	400 housing units and 1,500 jobs located 0.1 mile from Dallas Area Rapid Transit (DART) station	38
Montgomery county	Infill site near major transit center	42
San diego	Infill development project	52
West palm beach	Auto-dependent infill project	39

Source: From Center for Clean Air Policy (see Ref. 14).

- Car-share services within convenient walking distance (vehicle rental services designed to substitute for automobile ownership).
- Parking management (reduced parking requirements and renting parking spaces separately from building space, so residents who do not own an automobile are not forced to pay for parking they do not need).

BENEFITS AND COSTS

Location-efficient development can provide several benefits:

- Consumers benefit from more housing, shopping, and transportation choices, and from financial savings. Nondrivers in particular benefit from having housing options designed for maximum accessibility, as well as financial savings from reduced driving and parking costs. Per-household transportation expenditures tend to be lower for residents in such areas. Residents of cities with high levels of transit ridership tend to spend significantly less per capita on transportation than residents of more automobile-dependent cities,[2] as illustrated below. Similarly, McCann[18] found that households in more automobile-dependent communities on average spend more than 20% of their household budgets on transportation (more than $8,500 annually), whereas those in communities with more diverse transportation systems spend less than 17% (less than $5,500 annually), representing thousands of dollars in annual savings.
- By reducing per-capita vehicle ownership and use, location-efficient development tends to reduce per-capita traffic congestion and delays, road and parking facility costs, traffic crashes, pollution, energy consumption, and sprawl.
- By improving walking and cycling conditions, an increased portion of travel involves physical activity, which provides health benefits.
- Developers can benefit from having more design flexibility, including more opportunities for infill development and reduced parking costs. Also, LEMs increase the amount a household can spend on housing. It creates new markets and financing options. The New Urbanist movement promotes this type of development from an industry perspective.
- Regional economies tend to benefit when consumers shift their transportation expenditures from vehicles and fuel to transit services or general consumer goods.

There may be costs associated with higher population density in urban neighborhoods.[19] Higher density may increase congestion intensity (i.e., when people drive, they face greater congestion delay, although because they drive shorter distances and have alternative modes, they face less congestion delay per capita). Some households that choose location-efficient housing that has limited parking eventually may purchase additional motor vehicles if their needs change or they become wealthier, thus increasing local traffic and parking problems. This may require parking management. Some location-efficient housing includes resident covenants that restrict vehicle ownership. Urban infill may also cause displacement of lower-income households (gentrification).

EQUITY IMPACTS

Location-efficient housing and location-efficient mortgages tend to increase equity by allowing households that own fewer than average automobiles to avoid paying for parking they don't use, and by increasing housing options for lower-income households and nondrivers. Residential parking requirements reflect suburban, middle-class car ownership rates that are excessive for many households, particularly those with lower incomes. This is both unfair and regressive. Location-efficient development is optional, so consumers will choose it only if they consider themselves to be better off overall.

Location-efficient development and location-efficient mortgages tend to benefit lower-income households by providing financial savings and improving affordable transport and housing options.

Location-efficient development is most appropriate in urban neighborhoods that have good access (services and activities are easily available by walking and transit). It can be implemented by regional or local governments, or by not-for-profit organizations or individual businesses.

BEST PRACTICES

Here are some specific recommendations for implementing location-efficient development:

- It should include a variety of land-use and transportation features that improve access and mobility options, such as pedestrian and cycling improvements, transit improvements, and mixed land use.
- It should include a range of housing types and prices, so that people in various life-cycle stages and income classes can choose such housing.
- Parking requirements should be reduced or eliminated for location-efficient housing. Rather than being included with housing, parking should be rented separately, so that households pay only for the amount of parking they actually use.
- Parking should be managed to prevent spillover problems.

REFERENCES

1. Frank, L.; Kavage, S.; Litman, T. Promoting Public Health Through Smart Growth: Building Healthier Communities Through Transportation and Land Use Policies. Smart Growth BC (www.smartgrowth.bc.ca), 2006.
2. Litman, T. *Evaluating Public Transit Benefits and Costs*; VTPI (www.vtpi.org), 2005.
3. Frank, L.; Pivo, G. Impacts of mixed use and density on utilization of three modes of travel: SOV, transit and walking. Transportation Res. Rec. **1995**, *1466*, 44–55.
4. FHWA. Urbanized areas: selected characteristics. Highway Statistics 2004. U.S. Federal Highway Administration, (www.fhwa.dot.gov/policy/ohim/hs04/xls/htm72.xls), Table 72, 2005.
5. Moudon, A.V.; et al. *Effects of Site Design on Pedestrian Travel in Mixed Use, Medium-Density Environments*, Washington State Transportation Center, Document WA-RD 432.1, 1996. Available at: www.wsdot.wa.gov/ppsc/research/onepages/WA-RD4321.htm
6. Holtzclaw, J. Using Residential Patterns and Transit to Decrease Auto Dependence and Costs. National Resources Defense Council (www.nrdc.org), funded by the California Home Energy Efficiency Rating Systems, 1994.
7. Ewing, R.; Pendall, R.; Chen, D. Measuring Sprawl and its Impacts. Smart Growth America (www.smartgrowthamerica.org), 2002.
8. Lawton, K.T. The Urban Structure and Personal Travel: An Analysis of Portland, Oregon Data and Some National and International Data. E-Vision 2000 Conference (www.rand.org/scitech/stpi/Evision/Supplement/lawton.pdf), June 2001.
9. NPTS. National Personal Transportation Survey. U.S. Department of Transportation (www-cta.ornl.gov/cgi/npts), 1995.
10. Litman, T. *Land Use Impacts on Transportation*; VTPI (www.vtpi.org), 2006.
11. Arigoni, D. Affordable Housing and Smart Growth: Making the Connections. Subgroup on Affordable Housing, Smart Growth Network (www.smartgrowth.org), and National Neighborhood Coalition (www.neighborhoodcoalition.org), 2001.
12. Victoria Transport Policy Institute Online TDM Encyclopedia, 2005. Available at: www.vtpi.org.
13. CCAP. *Transportation Emissions Guidebook: Land Use, Transit & Travel Demand Management*; Center for Clean Air Policy, 2005. Available at: www.ccap.org/trans.htm
14. SGN. *Getting To Smart Growth: 100 Policies for Implementation*, and *Getting to Smart Growth II: 100 More Policies for Implementation*; Smart Growth Network (www.smartgrowth.org) and International City/Country Management Association (www.icma.org); 2002 and 2004.
15. Goldstein, D.B. Making Housing More Affordable by Correcting Misplaced Incentives in the Lending System. National Resources Defense Council (www.nrdc.org), 1996.
16. Hoeveler, K. Accessibility vs mobility: The location efficient mortgage. *Public Investment*. American Planning Association (www.planning.org) and Center for Neighborhood Technology (www.cnt.org/lem), 1997.
17. Russo, R. Planning for Residential Parking: A Guide For Housing Developers and Planners. Non-Profit Housing Association of Northern California (www.nonprofithousing.org) and the Berkeley Program on Housing and Urban Policy (http://urbanpolicy.berkeley.edu), 2001.
18. McCann, B. Driven to Spend. The Impact of Sprawl on Household Transportation Expenses, Surface Transportation Policy Project (www.transact.org), 2000.
19. Litman, T. *Evaluating Transportation Land Use Impacts*; VTPI (www.vtpi.org), 2004.

BIBLIOGRAPHY

1. Carman, E.; Bluestone, B.; White, E. Building on Our Heritage: A Housing Strategy for Smart Growth and Economic Development. Center for Urban and Regional Policy, Northwestern University (www.curp.neu.edu); October 2003.
2. Cervero, R. Built Environments and Mode Choice: Toward a Normative Framework. Transportation Res. D **2002**.
3. City of Orlando, Florida. Applicability of Vehicle Miles of Travel to Transporstation Planning, 1998. Available at: www.cityoforlando.net/planning/Transportation/ddc.htm
4. Ewing, R. *Transportation and Land Use Innovations; When You Can't Build Your Way Out of Congestion*; Planners Press: Chicago, IL, 1997; [www.planning.com].
5. Friedman, B.; Gordon, S.; Peers, J. Effect of neotraditional neighborhood design on travel characteristics. Transportation Res. Rec. **1995**, *1466*, 63–70.
6. Institute for Location Efficiency (www.locationefficiency.com) is an organization that works to encourage implementation of location-efficient development.
7. Jia, W.; Wachs, M. Parking Requirements and Housing Affordability: A Case Study of San Francisco. Research Paper 380, University of California Transportation Center (www.uctc.net), 1998.
8. LGC. Creating Great Neighborhoods: Density in Your Community. Local Government Commission (www.lgc.org), U.S. Environmental Protection Agency and the National Association of Realtors, 2004. Available at: www.lgc.org/freepub/PDF/Land_Use/reports/density_manual.pdf.
9. Litman, T. *Parking Requirement Impacts on Housing Affordability*; VTPI (www.vtpi.org), 2003.
10. Nelson, A.C. et al. The Link Between Growth Management and Housing Affordability: The Academic Evidence. Brookings Institution Center on Urban and Metropolitan Policy (www.brook.edu/dybdocroot/es/urban/publications/growthmang.pdf), 2002.
11. Nelson/Nygaard Consulting. Housing Shortage/Parking Surplus. Transportation and Land Use Coalition (www.transcoalition.org/southbay/housing_study/index.html), 2002.
12. Non-Profit Housing Association (www.nonprofithousing.org) provides a variety of materials to support development of affordable housing, including location-efficient development.
13. Reconnecting America. Hidden in Plain Sight: Capturing the Demand for Housing Near Transit. Center for

Ther–Und

Transit-Oriented Development; Reconnecting America; Federal Transit Administration (www.fta.dot.gov), 2004.
14. Regulatory Barriers Clearinghouse (www.huduser.org/rbc), created by the U.S. Department of Housing and Urban Development, was created to support state and local governments and other organizations seeking information about laws, regulations, and policies affecting the development, maintenance, improvement, availability, and cost of affordable housing.
15. Scheurer, J. Car-Free Housing in European Cities: A New Approach to Sustainable Residential Development. World Transport Policy Practice **1998**, *4* (3) [wwwistp.murdoch.edu.au/publications/projects/carfree/carfree.html].
16. The Affordable Housing Design Advisor Web site (www.designadvisor.org), sponsored by the U.S. Department of Housing and Urban Development, provides information on developing more affordable housing, redeveloping urban communities, and implementing Smart Growth.
17. The San Francisco Planning and Urban Research Association. Reducing Housing Costs by Rethinking Parking Requirements, 1998. Available at: www.spur.org.
18. West Coast Environmental Law Foundation. Smart Bylaws Guide, 2004. Available at: www.wcel.org/issues/urban/sbg.

Ther–Und

Tribal Land and Energy Efficiency

Jackie L. Francke
Geotechnika, Inc., Longmont, Colorado, U.S.A.

Sandra B. McCardell
Current-C Energy Systems, Inc., Mills, Wyoming, U.S.A.

Abstract

As energy service companies move beyond their traditional markets, technical competence must be combined with other skills, such as an understanding of different types of institutions and the ability to work with individuals of varying backgrounds and even cultures. In the United States, Tribal communities are one such market. This paper discusses important factors beyond engineering and economics that should be considered when working with Tribal communities, and presents an overview of experiences and recommendations for successful projects. Although the focus is on Tribal communities, the conclusions drawn are also applicable to rural communities and to projects in countries other than the United States—particularly those in developing countries where cultural traditions remain strong and energy service companies are as yet uncommon.

INTRODUCTION

The traditional markets for energy efficiency services are institutional customers, educational institutions of all kinds, and government entities—all institutions with large facilities having high utility bills, significant air handling requirements, and a reasonably uniform building configuration. As one moves outside these traditional markets to those that have historically been underserved, however, life can become both more complicated and more interesting for the energy service provider.

High on the list of underserved markets are facilities in developing countries and (inside the United States) rural and Tribal facilities, which provide challenges that are technical, operational, and above all cultural. To succeed in these types of projects, a company must include cultural considerations in its project implementation. The authors, with experience in implementing projects successfully in these varied environments, present projects on Tribal lands as a model for some of the approaches needed to accomplish energy efficiency programs in these areas.

As Tribal governments and communities gain and establish control of their natural resources, economies, and communities, they are faced with the challenges of developing and maintaining not only skilled labor pools, but also Tribal infrastructure and even sustainable economies. Often, all rural communities—and Tribal communities in particular—must perform and maintain necessary community services such as law enforcement, health services, and government functions. With the limited number of people in these areas, this can mean that opportunities for economic growth fall by the wayside.

Communications technologies and widespread travel are among the trends causing the world to shrink, so that communities that have until now been considered out of the mainstream are turning to urban resources and businesses to gain more knowledge and obtain services that will introduce them to new technologies, improve economic sustainability, and empower local government officials. For service companies working with these communities, it is important to remember that technology and engineering expertise may take a back seat to knowledge, understanding, and awareness of the community's culture. Working with Tribal governments and communities presents technical challenges, of course—but more important are the challenges of cultural differences and geographic location.

ENERGY EFFICIENCY ON TRIBAL LANDS

Background

There are more than 500 federally recognized Tribes in the United States. There are reservations in almost every state, and they range in size from less than one acre to many thousands of square miles, from cities to forests to beaches to deserts. The Tribes themselves are extremely varied, with different histories, cultures, languages, religions, and traditions. The United States recognizes the right of these Tribes to self-government and supports their Tribal sovereignty and self-determination. This means that each Federally recognized Tribe possesses the right to form its own government, to develop and enforce its own laws, to

Keywords: Energy efficiency; Tribal; Rural; Culture; Community.

tax, and exclude persons from Tribal territories, among other rights and responsibilities.

As a side note, it is often believed that, with the advent of Indian Gaming Laws that allow casinos on Tribal lands, all Tribes have adopted this approach to economic development and, therefore, are highly developed and profitable enterprises in themselves. But like all generalizations, this one is incorrect. Many Tribes have resisted gaming for cultural or other reasons, and even where gaming is present, the methods of distributing economic benefits vary widely.

In 2005 and 2006, a series of energy efficiency audits was performed on residential and commercial buildings located on Tribal lands. Under a U.S. Department of Energy grant directed by the Council for Energy Resource Tribes (CERT) based in Denver, Colorado, the purposes of the energy audits were to identify energy conservation opportunities for the Tribes and to establish alliances between teams of private experts and Tribes to develop energy efficiency.

The energy audits were performed on Tribal lands located in California, New Mexico, Oklahoma, and South Dakota by a team consisting of Indian and non-Indian personnel. The team traveled to the Tribal communities and worked closely with Tribal leaders and community members to identify areas that would improve the energy efficiency of the buildings. The results and recommendations were summarized in each case in a report, with copies submitted to Tribal leaders for future implementation.

The recommendations and scenarios discussed in this paper are based on the observations and experiences drawn from the performance of these energy audits on Tribal lands. The challenges the Tribal communities face with regard to geographic location, access to information, lack of access to subject-matter experts, and cultural awareness are similar to those faced by any small rural community in the United States—although the cultural gap can be wider (Fig. 1).

Fig. 1 Cultural and geographic challenges can be enormous in the open rural areas of the West. Where distances are so vast, the sense of community becomes doubly important.
Source: From Current-C Energy Systems, Inc.

Tribal Governments

Like many small rural communities, Tribal communities are often too small to be considered cities. Instead, as small towns and villages, they usually entrust government and community decisions to an elected board, council, or chapter, which may be made up of commissioners, community elders, or volunteers. Frequently, as in small towns, the positions are part time (though with full-time responsibility), and usually, a president, chairperson, or mayor functions as the chief executive officer. Naturally, in smaller communities such as these, the holding of political office is not a career path, but something one does on the side, as an obligation to the community or as a family responsibility, after achieving a good reputation in another field (Fig. 2).

With such a small governing council or board managing a small community where few resources are available, and with both limited funding and time, the primary goal of the board usually is meeting the most critical needs of the community. These include water supply, law enforcement, fire protection, health services, and the school system. Therefore, when opportunities arise for changes that may bring economic growth (energy efficiency improvements, capacity building in the energy arena, and renewable energy projects, for example), the community may have difficulty assessing the opportunities. The first issue here is the lack of a base of shared understanding such as that developed in larger cities by, for example, public relations and advertising campaigns; followed by a lack of skilled community experts, the

Fig. 2 In Tribal communities, people, places, and other components of the world are often given more respect than is true in the urban culture. "Shiprock" on the Navajo Reservation.
Source: From Geotechnika, Inc.

inadequacy of funding, and generally insufficient or inaccurate information. These community challenges require the Tribes to rely on outside resources to fill these voids and provide community opportunities. Evidently, therefore, there is a great possibility for service providers to work with Tribal communities—but it is important first to learn to think outside the traditional "energy service company" box. When working with Tribal communities, businesses must first gain a sense of understanding of the culture, the community and its priorities, and the way that particular government works. Only in this way will any project be successful and lead to future work in the community.

Observations

Cultural Awareness

As noted in the Background section earlier in this chapter, each Tribal community is different, with its own cultural values and traditions. In the context of developing and implementing a project, these differences in turn lead to differences in decision-making, leadership styles, and other business-related characteristics. It is important to understand that what may be observed in one village may not be valued in another. One must observe, learn, respect, and acknowledge the customs and values of each Tribal community.

A few general comments can be made, however, to highlight the types of situations and sensitivities of which any company operating on Tribal lands should be aware. These are not meant to be typical, as there is no such thing, but they are indicative.

It is not uncommon, for example, when working with Tribal communities that meetings be opened with a blessing/prayer led by a community leader or elder before any introductions and certainly before the business part of the meeting can start. When introductions begin, the focus is not on one's own individual accomplishments, but on one's family, heritage, and connection to the community. This stems from the importance of humility in many Indian cultures. Each person is considered to be equal to the others and no better. Therefore, in working with small Tribal communities, it is important to be humble and to focus importance on the Tribal elders, council members, and community members. Those who present a sense of their own importance or take themselves too seriously may violate Tribal cultural values and might be considered untrustworthy by the community.

Another common issue for those who have not worked with Tribal communities is that they must also understand the culture's ways of showing respect for another person. If, during a presentation to the community, the presenter does not see the direct eye contact to which he or she is accustomed, which in urban business culture usually means that the audience is involved, he or she should not feel as though the audience is disinterested in the topic at hand. In many Native American cultures, indirect eye contact may be considered a sign of respect.

Probably the most misunderstood value of American Indian culture (and the cultures of many developing countries) is the meaning of time. For one working with Tribal communities, patience and understanding of the meaning of time are of vital importance. In many American Indian cultures, time is not measured by the clock, but by when the situation feels appropriate. In many instances, Tribal community meetings may run late into the night or may not begin until all appropriate people are present. As in any small community where relationships are more important than deadlines, the necessity of dealing with daily issues may delay Tribal leaders; their commitment to the community is paramount, however, and resolutions of any local issues must be addressed. Everywhere in the world, the rhythms of small-community life are different—not necessarily more relaxed, as the workloads of these local leaders often put those of corporate chiefs to shame—and have a different focus. It is the community that is important. Therefore, service providers should be prepared to wait, if need be, and to change their schedules based on unforeseen circumstances. On the other hand, they should not take this as an excuse to be late themselves; that would be a strong sign of disrespect.

While working on Tribal lands, it is not uncommon to encounter elders and/or Tribal members who speak their native language as their primary language. While performing the energy audits, we had the opportunity to perform an energy audit on the home of an elderly Tribal couple not fluent in English. As in all the audits, a preaudit interview with the occupants was performed to gain a better understanding of energy usage in the home. During the interview, an interpreter was utilized. This situation also provided the team the opportunity to learn a few words in the local language. Most important was "Thank you"—a phrase that was returned with a smile at the end of the interview. Such opportunities should be appreciated, as they provide insight into the culture, show interest and respect, and provide to energy efficiency work a spice and flavor that are not often found in an office building or a hospital.

These examples are neither typical nor exhaustive, of course; with so many different cultures and so many different communities, it would be impossible for them to be so. If service providers are interested, genuine, respectful, and quietly professional, and if they watch and listen for cues from those around them, cultural awareness will develop.

Sense of Community

Business culture is often modeled on the culture of a big city—fast-paced, goal-oriented, with one-dimensional relationships and a separation between professional and personal life. Rural communities, including Tribal ones, are the centers of small-town life—even now, in the 21st century. Community life is important; relationships are of long standing; decisions are often made for reasons that are difficult for those from outside the community to understand; and discussions often appear to go back generations.

To work in such places, this sense of community must be understood and appreciated, and any projects must enhance that community to be successful. In addition, the traditional small-town value of "My word is my bond" is alive and strong in small towns and Tribal lands. When working in a community that has a shared memory going back for generations, it is important for service providers to say what they will do—and do what they say. Trust, which develops slowly in communities that are inherently wary of outsiders, is critical for success.

The clan system prevalent in the American Indian culture makes the community, in essence, a circle of family. This circle provides community members a sense of family support and a strong set of relationships. When accepted into this family, an outside service provider may be expected to assist the community in areas outside the project objectives, such as locating sources of information for the Tribal communities. In a community, boundaries are fluid, and when one joins the community, those fluid boundaries surround everyone.

Long-Term Orientation

In Tribal communities, an issue related to both the sense of community and the sense of time mentioned earlier is the long-term orientation taken to decision-making and project implementation. For most service providers, the process of proposing a project, having it accepted, and then implementing it is most familiar—and preferred. In Tribal communities, that process is often broken down into multiple steps on a time scale that is defined by other constraints, such as the needs of the community, Tribal leaders' lack of time, or insufficient funding. When a Tribal leader agrees that it is a good idea to implement a project, it is tempting to assume that he or she means now, or at least within the next few months. Unfortunately, the real meaning may be (as mentioned earlier) "when the time is right"—by which time the service provider may have given up and moved on to other projects.

It is important, therefore, for each party to understand the other's time scales and the steps that must be taken to move the project along one step at a time (with the associated approximate costs of those steps).

Geographic Location and Coordination

With an audit team of six, transporting equipment and personnel to each Tribal community required significant planning and preparation for each energy audit to be effective, successful, and within budget guidelines. On average, travel entailed one full day of travel by car to reach the sites, with three days on-site performing the energy audits.

To ensure that each trip was successful, documentation on energy usage pertaining to each building was requested in advance; meetings were scheduled; and a Tribal community point of contact was established, most often a Tribal leader. Because time runs differently in most American Indian cultures, this preparation required persistence and patience, as well as understanding. The possibility presented by using energy efficiency programs to bring new technology and opportunities to the community may be overridden by other daily priorities. We discovered that where possible, it was good to include the possibility of staying one extra day, just in case the schedule was changed during the course of the audit. (We always used that extra time).

Unlike most cities, where business representatives have the capability to travel to and from clients and conduct business in one day, rural areas and Tribal locations usually must be the single focus for that day or for several days. Therefore, additional planning, patience, and persistence play an important role (Fig. 3).

Technical Issues

In a standard commercial building, a school, or a hospital, the contact person for the energy audit is likely to be an engineer, an energy manager, or a technician familiar with the equipment in that particular building. In Tribal

Fig. 3 In rural and Tribal areas, a "community" often consists of a very few buildings in large land areas.
Source: From Current-C Energy Systems, Inc.

communities, there certainly are some such individuals, but it is the norm that those responsible for energy issues, construction, and maintenance also have significant other responsibilities. Given that situation, it is simply not possible for these individuals—no matter how dedicated and professional they may be—to be familiar with the opportunities provided by energy efficiency audits and enhancements. Neither is it often appropriate to recommend the latest, fanciest equipment unless the staff will be effectively trained on it and unless it can be serviced throughout its life.

As engineers, we are trained to recommend the best alternatives. On Tribal lands and in rural communities, however, the best technical solution may not be the best all-around solution. Cultural concerns, experience levels, power availability and reliability, time available for maintenance, training needs, service availability, and many other such factors must be taken into consideration for successful project implementation.

During our audits, we found that the most useful information often came from informal discussions, rather than from our carefully constructed audit forms and questionnaires. We also found that Tribal expertise resided in surprising areas (because many Tribal officials and personnel perform multiple roles), and in our implementation plans, we worked to use that expertise for training, project management, and other activities.

Access to Information and Administrative Notes

Unlike most urban communities, small Tribal communities may not have access to digital Internet services, but some may utilize dialup technology. Because of this, additional effort and time should be allocated in transferring information to the Tribal communities. Paper copies may need to be transmitted by mail or fax, depending on the content.

With this limited Internet access, most small communities are at a disadvantage when conducting research on such mainstays of business life as project opportunities and funding information. As an outside resource, a service business may add value by becoming the portal to that Internet-based information for Tribal communities.

During the energy audits, each visit was initiated with an on-site meeting. The meeting was conducted not only to build interest in and commitment to the audit, but also to convey and disseminate information regarding the energy efficiency audits and any new technologies available to the community. Although information was initially transmitted during the pretrip planning, in some instances, the point of contact did not receive the information or did not have the opportunity to review it due to other community concerns of higher priority. In addition, the meeting provided the community leaders the opportunity to discuss concerns and to question auditors on areas that they might be able to support during their visits (Fig. 4).

Therefore, a business expecting to work with Tribal communities should not rely on Internet access and should be understanding of Tribal leaders who may not consider the project to be a priority. Because of the limited time and higher priorities that come with being a Tribal leader, transferred information should be summarized and outlined, followed by contact with the Tribal officials to ensure that any questions, comments, or concerns he or she may have are addressed.

Funding Differences

With a typical energy service project in a city or a town, it is easy to determine who owns the facility, who is responsible for its construction and maintenance, and who would derive benefit from any energy efficiency measures installed. With buildings on Tribal lands, however, the picture is not so clear.

Ownership is a concept that is perceived differently among Tribes, with shared ownership by the entire Tribe, a family, or another group being the norm rather than the exception. In addition, because the laws on Tribal lands are often different from those off the reservations, banks and other financing institutions have historically been uncomfortable with the collateral or the foreclosure procedures required. The result has been that many facilities on reservations and other Tribal lands are inadequate and in

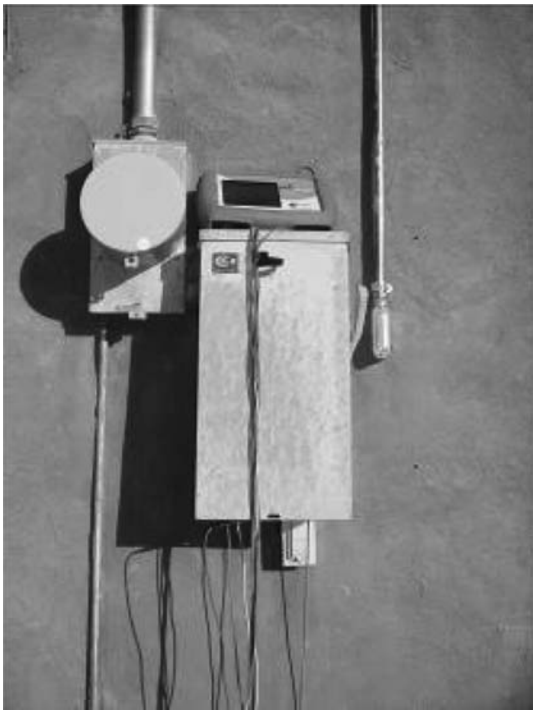


Fig. 4 Tradition meets technology: a power-quality meter determines that the wiring in a traditional adobe structure is inadequate and that the transformer is likely undersized. Source: From Current-C Energy Systems, Inc.

need of substantive repair or replacement. From an energy efficiency point of view, that presents an opportunity—if funds can be secured.

Often, the first place to look for funds that can be used to implement energy efficiency projects on Tribal lands is the federal government, which under its trust responsibility and as part of its current mandate is encouraging such projects. In addition, several states are focusing on increasing energy efficiency on the Tribal lands within their borders.

There are other instances in which funds are available for either energy efficiency or renewable energy programs from internal Tribal funds produced by casinos or other economic development activities, but government grants and loans form the largest proportion of funds available for such projects. It is important, therefore, for any service provider interested in working with Tribal entities to be familiar with the process of obtaining these grants to move projects forward.

Teamwork and Humility

If one were to develop an equation for the proportion of a successful Tribal project that is related to technical and engineering excellence, and the proportion that is attributable to the development of a project appropriate to the local culture, the importance of technical merit would likely be well under one-third of the project.

Service providers must understand, therefore, that in projects developed on Tribal lands or for rural areas (particularly those in developing countries), as possibly nowhere else, project design and development cannot succeed unless they are true team efforts, in which the service provider/consultant listens to and follows the lead of local Tribal leaders. Technical excellence on its own will go nowhere; only a project that meets the needs (technical, cultural, economic, environmental, and political) of the Tribe and its leaders, and that can be implemented within local constraints, can be accepted and successful.

For this team effort to be possible, service providers must leave the "expert" mantles they are accustomed to wearing in the urban business environment and work hard to be team members, catalysts, and learners—not experts. That requires a fair amount of both flexibility and humility.

One concrete example is that introductions should not focus on one's career accomplishments, but on one's family, heritage, and community. Focusing on one's accomplishments may be interpreted as showing that the presenter believes himself or herself to be better than the audience. In addition, initiating one's introduction with a welcome in the native Tribal language may be interpreted as a sign of respect but should be carried out only after relationships have been established. A sense of humility is important in becoming a part of the community.

Project Success

It should be evident that although the technology and economic opportunity in a recommended project may be unsurpassed, the success of any project on Tribal lands is very much dependent on a number of factors unrelated to engineering and economics. One must be persistent, humble, understanding, and giving (Fig. 5).

Tribal meetings should be initiated with protocols and processes to which community members are accustomed. This may include saying blessings/prayers, contacting a Tribal elder for sponsorship or introductions, working with Tribal leaders, and presenting introductions in a manner consistent with Tribal values. In addition, one must be prepared to present to the community meetings at a time deemed appropriate by the meeting coordinator, rather than at a time measured on the clock. This requires patience, as one may not be called on for a period of time.

Over time, when trust and relationships have been established, one becomes a part of the community. In this capacity, service providers then should be willing to support the community in areas that may not pertain to the project. In any small community, the lines that urban business culture draws around a project do not exist, and if one is part of a community, one helps as one can. Also, in most small communities that lack access to information, one may become a portal of information for the community by informing Tribal leaders of upcoming opportunities, new technologies, and additional information related to community needs or goals.

As stated previously, small Tribal communities are disadvantaged (in terms of access to services) by their locations. In working with these communities, one must be well prepared logistically and highly organized before

Fig. 5 Projects and technologies should be appropriate and meet needs, which may call for innovative and creative approaches. There are great renewable energy resources on Tribal lands, but their development sometimes requires different thinking about organizational structure, financing, management, and training.
Source: From Geotechnika, Inc.

starting a project; it is often not possible to just call up a supply house and ask it to make a delivery. For a project to proceed well, persistent (but polite) phone calls, faxes, and emails may be required to ensure that key players are available for support and that information has been transmitted to the appropriate contacts prior to on-site visits.

Because some small Tribal communities are tight–knit circles of families, the success of any project requires community involvement. The key players in the community may dictate the processes and goals of the project, but community Tribal members need to be motivated and informed of the benefits to themselves and to the community. Community members must feel that they are part of the decision-making process, which can be achieved by disseminating information at community meetings and events. If the community members are motivated, Tribal leaders will become interested and make the project a priority.

Fig. 6 Tradition and a sense of community are values that are becoming more respected in mainstream culture as well. In some Tribal areas where the cultural tradition is remaining in one place rather than practicing a nomadic lifestyle, traditional buildings are very well adapted to the climate and the area, and in fact are being studied by some architects as models for the architecture of the future.
Source: From Geotechnika, Inc.

CONCLUSION

As our audit teams discovered, the success of any project in a small Tribal community is very much dependent on factors other than engineering and economics.

While providing services to small Tribal communities, the authors learned much about the challenges that Tribal leaders and community members face on a daily basis. Location and access to information prove to be challenges when seeking out services required by the community. Services provided by outside organizations are critical to the viability of small Tribal communities, because they provide assistance in introducing new technologies and in fostering economic sustainability. This is particularly true in remote areas, because subject-matter experts may not be available in the area.

As a result, businesses seeking to work with small Tribal communities should be prepared to become part of the community, learning and understanding their values and customs. They should become culturally aware of the community in which they want to work, and they should learn what is important to the members and Tribal leaders (Fig. 6).

Service providers should understand the concerns and challenges of the community and support Tribal leaders by becoming a portal of information leading to interesting opportunities, new technologies, and other pertinent information. They should also empathize with community leaders and members by lending support where needed.

Overall, and most important, one must be persistent, patient, and humble for the project to succeed. As stated earlier, the technology and opportunity may be unsurpassed from the standpoint of engineering and economics, but if humility, understanding, empathy, and cultural awareness do not exist, neither will the project.

About the Authors

Jackie Francke is a Tribal member of the Navajo Nation Tribe, a mining engineer, and president and principal engineer of Geotechnika, Inc. Francke has more than 15 years of experience working in the field of instrumentation, monitoring, and data collection on projects related to energy efficiency, mining, civil, and environmental engineering. Having grown up on the reservation, Francke understands the challenges Tribes face on a daily basis, and hopes to educate and assist businesses that are looking to work with Tribal communities.

Sandra McCardell is president and founder of Current-C Energy Systems, a small woman-owned firm specializing in efficient and sustainable energy applications in institutional, industrial, and community settings. Her experience over 25 years has ranged from consumer-products companies to international development to real estate to engineering. With a strong background in business and significant experience working in other countries and cultures, she particularly enjoys her projects with Tribal entities.

Underfloor Air Distribution (UFAD)☆

Tom Webster
Fred Bauman
Center for the Built Environment, University of California, Berkeley, California, U.S.A.

Abstract
The purpose of this paper is to provide an overview of the principles, features, benefits, and limitations of the building conditioning technology called underfloor air distribution (UFAD) and the closely related task/ambient conditioning (TAC).

INTRODUCTION AND BACKGROUND

Recent trends in today's office environment make it increasingly more difficult for conventional centralized heating, ventilation, and air-conditioning (HVAC) systems to satisfy the environmental preferences of individual office workers using the standardized approach of providing a single uniform thermal and ventilation environment. Since its original introduction in West Germany during the 1950s, the open plan office, containing modular workstation furniture and partitions, is now the norm. Thermostatically controlled zones in open plan offices typically encompass relatively large numbers of workstations in which a diverse work population having a wide range of preferred temperatures must be accommodated. Modern office buildings are also being impacted by a large influx of heat-generating equipment (e.g., computers and printers) whose loads may vary considerably from workstation to workstation. Offices are often reconfigured during the building's lifetime to respond to changing tenant needs, affecting the distribution of within-space loads and the ventilation pathways among and over office partitions. Compounding this problem, there has been growing awareness of the importance of the comfort, health, and productivity of individual office workers, giving rise to an increased demand among employers and employees for a high-quality work environment.

Underfloor Air Distribution

In the 1970s, underfloor air distribution (UFAD) was introduced into office buildings in West Germany as a solution to cable management and heat load removal issues caused by the proliferation of electronic equipment throughout the office.[13] In these buildings, the comfort of the office workers had to be considered, giving rise to the development of occupant-controlled localized supply diffusers to provide task conditioning. Some of the first UFAD systems in Europe used a combination of desktop outlets for personal comfort control and floor diffusers for ambient space control.[12]

Prior to the 1990s, office installations using underfloor systems were found primarily in South Africa, Germany, and other parts of Europe. The technology was not commonly used in North America before about 1995, in part due to the downturn in office building construction beginning in the mid-1980s. Japan did not experience this same downturn, and as a result, significant growth in UFAD technology was observed during this period. Between 1987 and 1995, more than 250,000 m^2 (2.7 million ft^2) of office space in more than 90 buildings was installed with UFAD systems in Japan.[14]

However, in the late 1990s, growth for raised floor installations in the United States was dramatic, and designers and manufacturers predicted that 35% of new offices would use raised floors by 2004. Half of these installations were expected to incorporate UFAD technology. This rate of increase slowed in 2003 due to the economic downturn and reduced office construction, but has revived since then. Installation data shows that since 2005, 12% of new commercial offices used raised floors, and 45% of those projects used UFAD systems.[11] We estimate that as of 2006, at least 400 UFAD projects have been built in North America.

Task/Ambient Conditioning (TAC)

During recent years, an increasing amount of attention has been paid to air distribution systems that individually condition the immediate environments of office workers within their workstations to address the issues outlined above. As with task/ambient lighting systems, the controls for the task components of TAC systems are partially or

☆ A version of this report was originally published by the U.S. Department of Energy, Office of Federal Energy Management, under the title *Technology Focus: Alternative Air Conditioning Technologies: Underfloor Air Distribution (UFAD)*, DOE/EE-0295, 2004.

Keywords: Constant air volume; Stratification; Swirl diffuser; Task ambient condition; Underfloor air distribution; Underfloor plenum; Variable air volume.

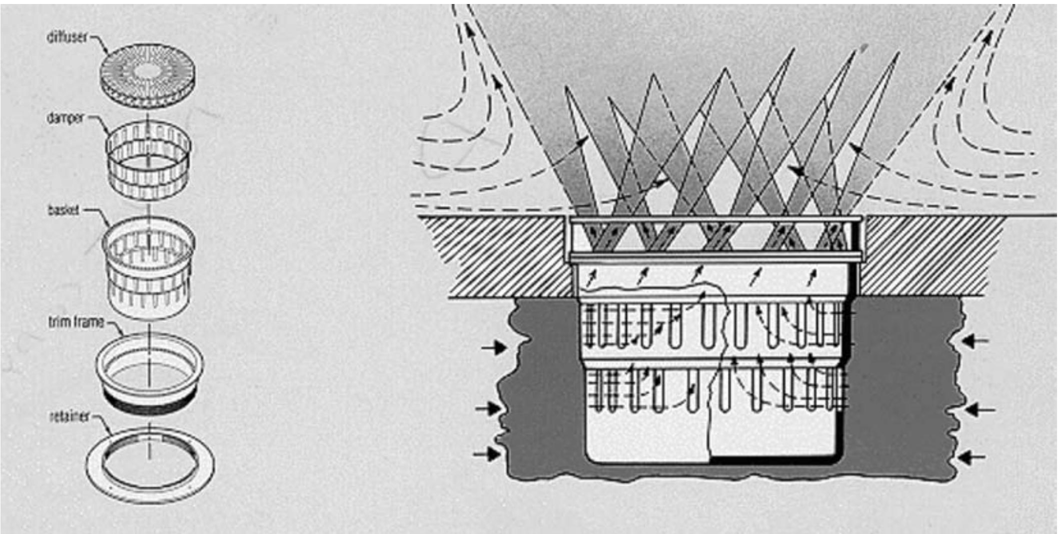


Fig. 1 Illustration of typical swirl diffuser used in underfloor air distribution (UFAD) systems.

entirely decentralized and under the control of the occupants. Typically, the occupant has control over the speed and direction, and in some cases the temperature, of the incoming air supply. Variously called "task/ambient conditioning," "localized thermal distribution," and "personalized air conditioning," these systems have been mostly installed in open-plan office buildings to which they provide supply air and (in some cases) radiant heating directly into workstations.

TECHNOLOGY DESCRIPTION

UFAD Systems

Underfloor air distribution systems use the open space (underfloor plenum, also used for distribution of other services such as power and computer cabling) between a structural slab and the underside of a raised floor system to deliver conditioned air to supply diffusers (Fig. 1) located at or near floor level within the occupied zone (up to 6-ft [1.8-m] height) of the space. The supply diffusers can provide some degree of individual control over the local thermal environment, depending on diffuser design and location. Additional supply diffusers provide ambient environmental control in nonwork areas. Active diffusers (for purposes of this paper) are defined as those with local means of volume adjustment (such as an integral variable speed fan or damper) that is amenable to automatic zone control (in addition to means for occupant control). Passive diffusers, although they may have means for occupant adjustment, are combined with terminal or system elements to achieve zone control. Systems designed with all fan-assisted active diffusers typically use zero-pressure plenums. Passive diffusers require pressurized plenums. The majority of UFAD systems currently being deployed have pressurized plenums with either active or passive diffusers.

TAC Systems

Task/ambient conditioning systems can be distinguished from standard UFAD systems by their higher degree of personal comfort control provided by the localized supply outlets. Task/ambient conditioning supply outlets use direct velocity cooling to achieve this level of control, and are therefore most commonly configured as fan-driven (active), jet-type diffusers that are part of the furniture or partitions or from floor outlets. A majority of TAC systems deployed have used UFAD to distribute conditioned air to fan-driven diffusers located in workstation furniture and partitions. Fig. 2 shows an illustration of such a TAC system.

For further information on a complete range of TAC systems, refer to Bauman et al. and Loftness et al.[6,10] Because few TAC systems are currently (as of 2006) being deployed, the remainder of this report will be focused only on UFAD systems.

PRINCIPLES OF OPERATION OF UFAD SYSTEMS

Figs. 3 and 4 illustrate the fundamental differences between traditional overhead and UFAD systems, respectively. As shown in Fig. 3, overhead systems (in office buildings, these are predominately variable air volume (VAV) all-air distribution systems) employ an extensive array of ductwork and terminal devices to provide supply air through the ceiling-mounted diffusers. Often referred to

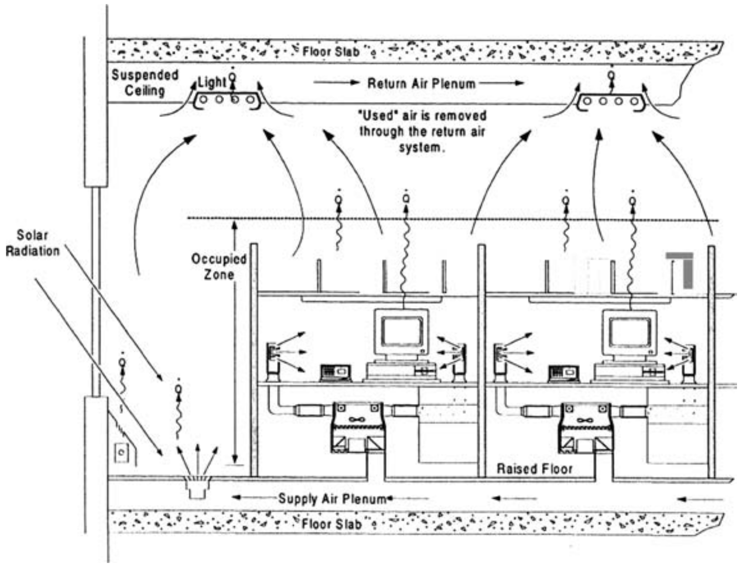


Fig. 2 Task/ambient conditioning (TAC) system.

as mixing ventilation systems, these systems are designed to promote complete mixing of supply air with room air, thereby maintaining the entire volume of air in the space at the desired temperature setpoint. Space air is typically returned to the AHU through an open ceiling plenum that also contains various other systems for lighting, electrical, communications, and fire protection.

Underfloor air distribution systems turn this concept upside down and have the following characteristics:

- Supply air, including at least the minimum required volume of outside air, is filtered and conditioned to the required temperature and humidity by a conventional AHU and passed through a minimum amount of ductwork to an underfloor plenum. The underfloor plenum is formed by installation of a raised floor system, typically consisting of $0.6\times0.6\ m^2$ ($2\times2\ ft^2$) of concrete-filled steel floor panels positioned 0.3–0.46 m^2 (12–18 in.) above the concrete structural slab of the building. The raised floor system also allows all cable services, such as power and communication, to be located in the plenum and provides easy access for modifications and maintenance.
- Individual office workers can control their local thermal environment to some extent (typically by adjusting the volume and/or trajectory of the supply air entering the space), giving them the opportunity to fine-tune the thermal conditions in their workstation to

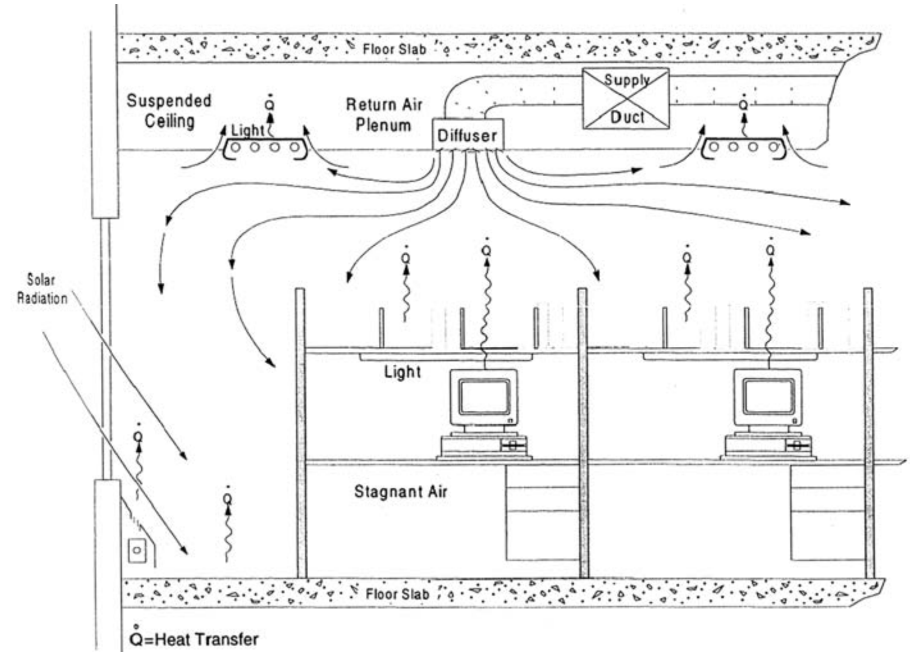


Fig. 3 Overhead system.

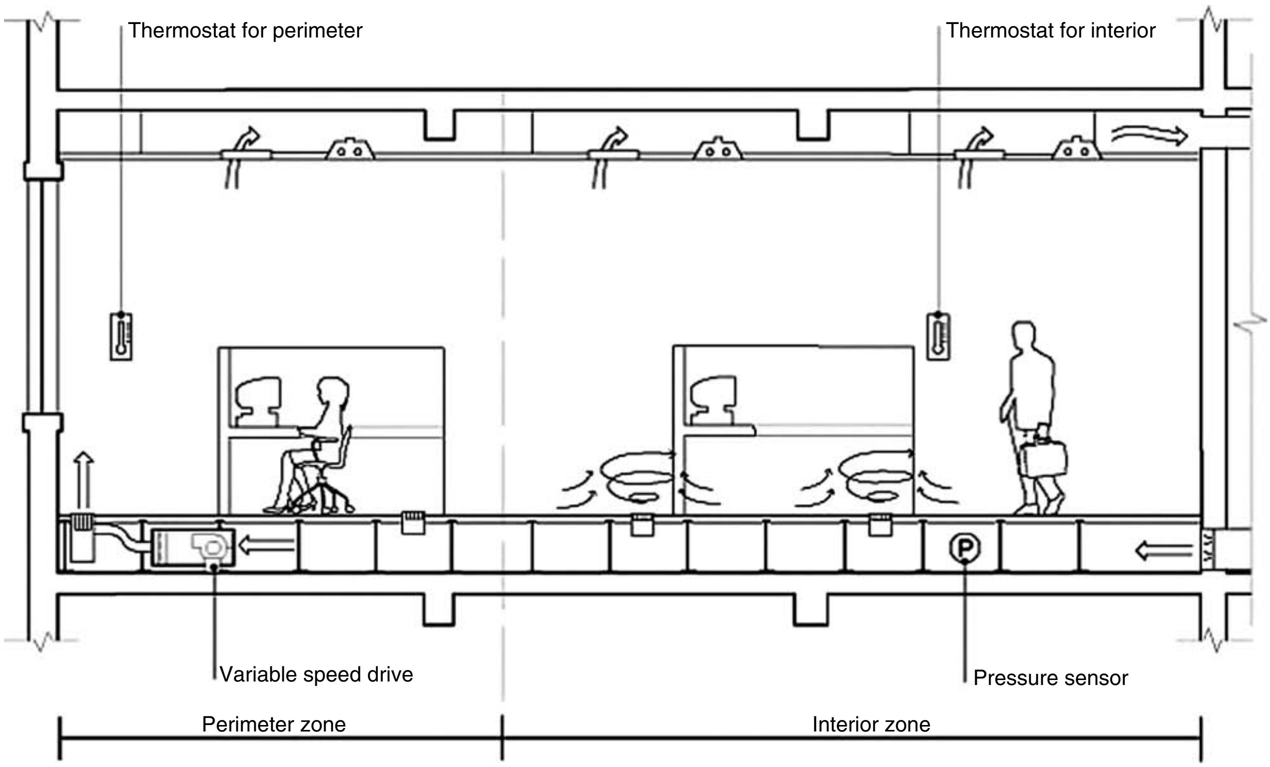


Fig. 4 Pressurized plenum underfloor air distribution (UFAD) system.

their personal comfort preferences. Different supply outlet configurations may be used depending on the conditioning requirements for a particular zone of the building, as discussed below.

- Underfloor air distribution systems benefit from a floor-to-ceiling airflow pattern that takes advantage of the natural buoyancy produced by heat sources in the office to efficiently remove heat loads and contaminants from the space. Air is returned from the room at ceiling level through recessed lighting fixtures and return grilles through a conventional ceiling return plenum or return grilles located high in the space when a return plenum is not used.
- Because the air is supplied directly into the occupied zone (up to 6-ft [1.8-m] height), supply outlet temperatures are generally maintained above 17°C–18°C (63°F–64°F) to avoid uncomfortably cool conditions for the nearby occupants and to minimize cool temperature near the floor.

There are a wide variety of approaches being used to provide a combination of individual and automatic zone control for UFAD systems.[2] The authors have identified over a dozen variations of these systems. Typically, these systems use VAV or constant-air-volume (CAV) methods for general zone control (i.e., overall zone control other than local occupant control).

- Six types of diffusers are currently being offered:
  — Fan-assisted, active
  — Variable area, active
  — Swirl, passive
  — Swirl, active
  — Linear bar grille, passive
  — Linear bar grille, active
- The heating and cooling loads of perimeter zones are handled by fan-powered constant volume or variable volume terminal reheat units located in the underfloor plenum (similar to those shown in Fig. 4). Passive or active diffusers are located in occupied areas of the zone, normally within about 4.5 m (15 ft) from the exterior walls. Passive diffusers are generally supplied by a (series) fan-powered mixing box or fan coil unit, or a VAV box either connected to the diffuser by ducting or by supplying air to a partitioned area of the plenum where the diffusers are located.
- Interior zones are generally large zones each controlled by one thermostat, but with diffusers located near the occupants within or close to their workstations.

Fig. 4 shows a schematic diagram and Fig. 5 shows a typical swirl diffuser layout for a pressurized plenum UFAD system.

This system is being commonly applied to office buildings due to its simplicity and cost savings. Although

Fig. 5 Typical interior swirl diffuser layout for underfloor air distribution (UFAD) systems.

this floor-based air distribution system provides somewhat limited individual comfort control for occupants, it still retains the same flexibility and potential energy-saving benefits associated with the others.

POTENTIAL OF UFAD APPLICATIONS

Benefits

Improved Thermal Comfort for Individual Occupants

Occupant thermal comfort is perhaps the area of greatest potential improvement in that UFAD systems potentially can accommodate individual differences. In today's work environment, there can be significant variations in individual comfort preferences due to differences in clothing and activity level (metabolic rate), as well as differences in the local heat gains and losses. By allowing personal control of the local thermal environment, UFAD systems could potentially satisfy virtually all occupants, including those out of thermal equilibrium with their surrounding ambient environment, as compared with the 80% satisfaction quota targeted in practice by existing thermal comfort standards.[1]

Improved Air Movement and Ventilation Effectiveness; Cleaner Environment

Some amount of improvement over conventional uniformly mixed systems is expected by delivering the fresh supply air near the occupant and at the floor.

Improved Occupant Satisfaction and Increased Worker Productivity

Underfloor air distribution systems have the potential to increase the satisfaction and productivity of occupants, because occupants have the ability to individually control their workspace environments. The financial implications of such improvements can be extremely large, as salary costs typically make up at least 90% of all costs (including construction, operation, and maintenance) over the lifetime of a building.

Energy Savings

Energy savings over conventional overhead systems are predominately associated with two factors: (1) cooling energy savings from economizer operation at higher supply air temperatures and (2) fan energy savings due to reduced static pressure requirements and potentially lower airflow requirements due to optimized stratification. Additional details of how these savings are achieved can be found in Bauman and Bauman and Lehrer.[2,3]

LIMITATIONS

Among the items that limit the widespread acceptance and application of UFAD technology are the following.

New and Unfamiliar Technology

For the majority of U.S. building owners, developers, architects, engineers, and equipment manufacturers, UFAD systems still represent a relatively new and unfamiliar technology. The decision to select a UFAD system will initially require changes in common practice, including new procedures and skills in the design, construction, and operation of such systems, as well as changes in responsibilities of the various installation trades. This situation creates some amount of perceived risk to designers and building owners. Significant progress has been made in these practices over the past several years.

Perceived Higher Costs

Many designers are concerned about the higher first costs of the raised flooring required for UFAD systems. However, as described above, there are many factors associated with raised floor systems that contribute to reducing life-cycle costs in comparison to traditional air distribution systems. (See the “Cost Effectiveness” section below). In a recent study of UFAD costs, we found that there are several ways that UFAD systems can be tailored to reduce the cost differential. Furthermore, the difference depends significantly on the quality of the traditional VAV system being used for comparison.[16]

Limited Applicability to Retrofits and Certain Building Types/Areas

The installation of UFAD systems and the advantages that they offer are most easily achieved in new construction. Some of the key system features are not always suitable for retrofit applications (e.g., access floors cannot be installed in existing buildings with limited floor-to-floor heights). Although widely applicable, there are some building types and areas within buildings where access floors and UFAD are not appropriate. These areas are generally those in which spillage has the potential to occur, including bathrooms, laboratories, cafeterias, and shop areas.

Lack of Information, Design Guidelines, and Evaluation Methods

Although in recent years there has been an increased number of publications on UFAD technology, only very recently has there been effort aimed at providing a complete understanding of the underlying fundamental fluid mechanics and thermal issues to allow for the creation of design and simulation tools. New guidelines and research results are presented in several sources.[4,5,9,15,17] System commissioning procedures, operating sequences, and control techniques are likewise under development.

Potential for Higher Building Energy Use

As with any space conditioning system, a poorly designed and operated UFAD system has the potential to use more energy than that of a well-designed conventional system. Other factors that can influence the energy comparison with conventional systems are number of perimeter fan-powered units, supply plenum thermal decay in UFAD systems, and airflow requirements.

Thermal Discomfort

Underfloor air distribution systems are perceived by some to produce cold floor, and because of the close proximity of supply outlets to the occupants, the increased possibility of excessive draft exists. However, we are finding that thermal comfort problems are more likely to be associated with inadequate control and operating strategies and inexperience of operators with UFAD technology.

Plenum Leakage

There have been reports of problems with leakage from the supply plenums either into the space (through floor penetrations and floor panel leakage) or through bypassing the occupied space to the return or outside the building altogether. This latter type of leakage, if excessive, is a serious problem, because it causes increased fan energy use. This is probably the single most serious problem found upon start-up of UFAD projects to date and can be traced to poor design and construction practices related to the leakage sources.

Virtually all of the issues listed above are actively being researched (refer to References) or addressed by design and construction professionals and equipment vendors in response to market demand. Some issues suggested by critics of UFAD technology have turned out not to be problems, and many of the problems that do occur can be traced to inadequate design, construction, and operating practices that have resulted from lack of knowledge and experience. Problems like these are typical for early installations of any nascent technology.

COST EFFECTIVENESS

Cost considerations will be different depending on whether the installation represents new or retrofit construction. Total first costs (shell and core plus tenant improvement) for UFAD systems using raised flooring will likely be somewhat higher than those for a conventional system. Preliminary results from research studies[16] have shown total building first costs of pressurized UFAD systems in new construction (including raised floor and structural differences) to be about $3.50 per gross square foot (gsf) greater than typical practice conventional VAV systems. However, if a raised floor system has already been justified for other reasons, such as improved cable management, or a high quality conventional system is the standard of comparison, the cost premium is effectively eliminated altogether, and under some circumstances, there can be savings in the range ∼$2–7/gsf. In new construction, UFAD can lead to reducing floor-to-floor heights, thus reducing structural costs. Furniture-based and active diffuser-based systems will generally cost more than other solutions.

Operating costs can be reduced in accordance with the energy-saving strategies discussed above. With the improved thermal comfort and individual control provided by UFAD systems, occupant complaints requiring

Table 1 Example churn rates

Group	Churn (%)	Office plan (% office/ open/bullpen)
Services	37	35/55/10
Manufacturing	40	34/58/8
Institutional/ fovernment	23	67/20/13

response by facility staff can be minimized. Underfloor air distribution systems using raised flooring provide maximum flexibility and significantly lower costs associated with reconfiguring building services (when changes are being made in the office layout) due to churn, and thus reduce life-cycle costs substantially.

First cost for retrofits, generally the bulk of construction activity, is most likely greater than those for new construction. Also, as indicated previously, some facilities are not amenable to retrofit by UFAD systems.

In order to determine total life-cycle cost differences between UFAD systems and conventional designs, the following factors must be considered in addition to first cost:

- *Churn*. This is the cost associated with relocating personnel and it is defined as the ratio of total workplace moves in a year to the total number of building occupants. These figures vary widely by industry type and building activities. Results of a study by IFMA[8] are shown in Table 1.

  As shown, government facilities have significantly lower churn than other industries, which is also reflected in the much lower percentage of open plan space. This indicates that the benefits of reduced churn costs in federal facilities may be limited. It should be noted that the cost of churn can vary considerably depending on the extent of the reconfiguration; i.e., simply moving to a new cubicle is much different that reconfiguring the layout of cubicles or offices. The high proportion of private offices in government facilities could drive these costs significantly if the reconfiguration involves more than office to office moves.
- *Operations and maintenance (O&M)*. This item includes the costs of maintenance and repair as well as energy. While energy savings estimates are limited due to the lack of appropriate capabilities in energy simulation programs and data from monitored projects, indications are that savings in annual HVAC system energy can be in the range of 0%–20% depending on system design and weather conditions. Maintenance costs are expected to be less than conventional systems due to the ease of access to the distribution system. However, commissioning/startup costs may be greater because the location and operation of the diffusers may require fine-tuning to optimize the occupant interaction benefits.
- *Productivity and health*. The savings associated with productivity and health benefits are difficult to measure and require considerably more research. However, recent studies[7] indicate that work performance improvements of 0.5%–5% may be possible if the indoor environmental quality is improved.

SUMMARY AND CONCLUSIONS

Underfloor air distribution systems have significant potential advantages compared with traditional VAV systems. Rarely has there been a space conditioning technology that promises the combined benefits of improvements in thermal comfort, energy efficiency, and productivity and health. Underfloor air distribution technology, like all nascent technologies, is being advanced both in theory and practice by researchers, designers, manufacturers, and early adopter owners who are working to bring the design, operation, and costs to the point where they can be more easily and reliably applied. Underfloor air distribution technology may someday displace overhead VAV as the system of choice for space conditioning. Although the use of UFAD systems in particular is becoming more common in the private commercial sector, the overall potential for UFAD in federal facilities may be limited by low churn that reduces life-cycle cost benefits. In addition, the overall federal building stock is not as amenable to UFAD installations as the private sector, due to the higher cost of retrofits (i.e., a greater ratio of fixed private offices).

However, in those situations where it is appropriate, there are many compelling reasons to consider UFAD for the space conditioning solution.

ACKNOWLEDGMENTS

Much of the material for this report was derived from work done at the Center for the Built Environment (CBE) at the University of California, Berkeley. The authors would like to gratefully acknowledge the support CBE and its industry partners, CBE sponsor the National Science Foundation (NSF), the Regents of the University of California, and the DOE/FEMP New Technology Development Program for providing funding for this report. CBE's current Industry Partners include: Armstrong World Industries; Arup; California Department of General Services; California Energy Commission; Charles M. Salter Associates Inc.; Flack & Kurtz Inc.; HOK, Pacific Gas & Electric Company; Price Industries; RTKL Associates Inc.; Skidmore; Owings & Merrill, LLP;

Stantec; Steelcase, Inc.; Syska Hennessy Group; Tate Access Floors, Inc.; Taylor Team: CTG Energetics, Guttmann & Blaevoet, Southland Industries, Swinerton Builders, Taylor Engineering; Trane; U.S. Department of Energy; U.S. General Services Administration; Webcor Builders; York International Corporation.

REFERENCES

1. ASHRAE. ANSI/ASHRAE Standard 55-2004. In *Thermal Environmental Conditions for Human Occupancy*; American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc.: Atlanta, 2004.
2. Bauman, F. *Underfloor Air Distribution (UFAD) Design Guide*; ASHRAE, American Society of Heating, Refrigerating, and Air-Conditioning Engineers: Atlanta, 2003.
3. Bauman, F.; Lehrer. D., [Underfloor Technology Website]. Center for the Built Environment (CBE), Available from: http://www.cbe.berkeley.edu/underfloorair/, 2006.
4. Bauman, F.; Jin, H.; Webster, T. Heat transfer pathways in underfloor air distribution (UFAD) systems. ASHRAE Trans. **2006**, *112* (Part 2).
5. Bauman, F.; Webster, T.; Jin, H. Design guidelines for underfloor air supply plenums. HPAC Eng. **2006**, *July*.
6. Bauman, F.S.; Arens, E.A. *Task/Ambient Conditioning Systems: Engineering and Application Guidelines*; Center for Environmental Design Research, University of California: Berkeley, 1996.
7. Fisk, W.J. Health and Productivity Gains from Better Indoor Environments and Their Relationship with Building Energy Efficiency, LBNL-45484, Lawrence Berkeley National Laboratory, July 31, 2000.
8. IFMA. *Benchmarking*; International Facilities Management Association: Atlanta, GA, 1997.
9. Jin, H.; Bauman, F.; Webster, T. Testing and modeling of underfloor air supply plenums. ASHRAE Trans. **2006**, *112* (Part 2).
10. Loftness, V.; Brahme, R.; Mondazzi, M.; Vineyard, E.; MacDonald, M. Energy savings potential of flexible and adaptive HVAC distribution systems for office buildings—Final report. Air Conditioning and Refrigeration Technology Institute 21-CR Research Project 605-30030, June. Full report available at www.arti-21cr.org/research/completed/index.html, 2002.
11. Reynolds, B. Personal communications, Product manager, Tate Access Floors, Inc.: Jessup, MD, 2006.
12. Sodec, F. Air Distribution Systems Report No. 3554A, Krantz GmbH & Co.: Aachen, West Germany, September 19, 1984.
13. Sodec, F.; Craig, R. The underfloor air supply system—the European experience. ASHRAE Trans. **1990**, *96* (Part 2).
14. Tanabe, S. *Task/Ambient Conditioning Systems in Japan*, presented at Workshop on task/ambient conditioning systems in commercial buildings, San Francisco, May 4–5, 1995.
15. Webster, T.; Bauman, F. Design guidelines for stratification in underfloor air distribution (UFAD) systems. HPAC Eng. **2006**, *June*.
16. Webster, T.; Benedek, C.; Bauman, F. *Underfloor Air Distributsion (UFAD) Cost Study: Analysis of First Cost Tradeoffs in UFAD Systems*, Center for the Built Environment, University of California, Berkeley, September 2006.
17. Webster, T.; Bauman, F.; Reese, J.; Shi, M. *Thermal Stratification Performance of Underfloor Air Distribution (UFAD) Systems*, Proceedings, Indoor Air 2002, Monterey, CA, June 2002.

BIBLIOGRAPHY

1. Bauman, F.; Webster, T. Outlook for underfloor air distribution. ASHRAE J. **2001**, *43* (6), 18–27.

Utilities and Energy Suppliers: Bill Analysis☆

Neil Petchers
NORESCO Energy Consulting, Shelton, Connecticut, U.S.A.

Abstract
Once the energy rate structures for a customer have been examined and been understood, the next step in understanding how utility costs are determined is to perform a utility bill analysis. Knowing how a customer is charged for the energy it uses each month is an important piece of the overall process of energy management at a facility. This article discusses how electric bills and gas bills for large commercial, industrial, and institutional customers are calculated. Average costs, peak load costs, and time-of-use (TOU) costs are presented and evaluated. Electric costs are found as demand costs per kilowatt per month, and energy costs per kilowatt hour. The utility cost data are used initially to analyze potential energy savings opportunities, and will ultimately influence which of these opportunities are recommended.

INTRODUCTION

The ability to calculate utility bill impacts is fundamental to the evaluation of any energy system. An effective method for determining the energy operating cost impact of a proposed option is to compare the calculated annual utility bills with and without the option in place. The difference represents the energy operating cost impact associated with the option. To do so, one must carefully identify not only the aggregate change in fuel or electricity usage associated with the proposed option, but when the changes occur with respect to the various time periods and other billing factors integral to the rate schedule.

The average unit cost for utilities defined as the total annual cost divided by the annual consumption, is a tempting simplification of utility rate structures. It allows easy calculation of energy cost impact from changes in energy usage patterns. In some cases, this simplistic approach yields accurate answers. More typically, however, application of average unit costs gives results that are inaccurate and can, at times, be very misleading.

For any proposed technology application that would change the number of fuel or electricity units consumed, the cost impact can rarely be accurately determined by merely multiplying the change in units by the average cost per unit for the total facility usage. This is because incremental costs are usually quite different from average costs. With most currently available utility rates, the addition or subtraction of usage during peak periods may have several times the cost impact as the addition or subtraction of the same amount of usage in off-peak periods. Therefore, one must consider the weighted average cost of the increase or decrease in consumption units associated with a proposed application.

CALCULATING UTILITY BILLS

Utility bills can typically be broken down into the following basic components:

- A customer or minimum service charge
- An energy or commodity charge
- A demand or maximum level of service charge
- Power factor penalties
- Adjustments such as taxes levied by state, county, and city authorities
- Surcharges or credits associated with specific orders established by various regulatory authorities
- Fuel cost adjustments that reconcile the actual cost of fuel used or delivered by the utility, with the estimated cost used in the most recent rate proceeding to set the energy or commodity charge.

These basic components are expressed and calculated in many ways using various units of measure. Combined, they comprise the total utility bill with each contributing in different ways to the weighted average cost. To calculate a utility bill, one must carefully read the rate tariff inclusive of all rate riders and adjustment clauses. One must also know the current values for items that vary such as fuel adjustment charges. Based on this information, one should be able to calculate the utility bill exactly. If such calculations

☆ This entry originally appeared as "Utility Bill Analysis" (Chapter 22) in *Combined Heating, Cooling & Power Handbook*, by Neil Petchers. Published by Marcel Dekker, Inc., New York, NY, 2003; reprinted with permission from AEE/Fairmont Press.

Keywords: Utility bill analysis; Utility cost analysis; Utility bills; Electric bills; Gas bills; Electric costs; Gas costs; Energy bills; Energy costs; Average cost of power.

do not equal the utility bill exactly, either a mistake has been made in the computation or a piece of information is missing.

Utility rate spreadsheets are useful in performing such computations. Once a spreadsheet is built, it can be used to calculate costs for any usage pattern under a given rate or set of rates. It can also be used to quickly calculate cost savings from energy efficiency improvements.

TYPICAL GAS BILL CALCULATION

The following is a sample utility bill calculation for a given natural gas usage profile for natural gas service. This is a typical declining block rate structure, where different levels of usage are billed at different unit costs. In this case, the billing unit is 100 cubic feet (Ccf) of natural gas. Hundred cubic feet is a commonly used volume of gas for billing purposes. Other commonly used billing units are 1000 cubic feet (Mcf), therm (100,000 Btu), and million Btu (MMBtu). The rate schedule is shown in Fig. 1. Refer to Chapter 5 for details on the energy or heat content of natural gas billing units. Assuming that the gas usage for the month was 47,500 Ccf, the bill would be calculated as follows:

The commodity rate is first adjusted to account for the purchased gas adjustment (PGA) and the demand side management (DSM) surcharge. To each rate block, \$0.0097 is subtracted to account for the PGA and \$0.0020 is added to the commodity rate to account for the DSM surcharge. The resulting net commodity rate is:

Block	Ccf	Per Ccf
First	10,000	\$0.4297
Next	20,000	\$0.4145
All over	30,000	\$0.4023

Monthly service charge: \$80.00

Minimum monthly charge = The service charge

Commodity rate:

Block	**Ccf**	**Per Ccf**
First	10,000	\$0.4374
Next	20,000	\$0.4222
All Over	30,000	\$0.4100

Purchased gas adjustment (PGA):	(\$0.0097)
Demand-side management (DSM) charge:	\$0.0020

State tax: 4.3%
City tax: 1.5%

Fig. 1 Sample natural gas rate tariff.

Therefore,

For the first 10,000 Ccf, the charge is:	\$4,297.00
For the next 20,000 Ccf, the charge is:	\$8,290.00
For the final 17,500 Ccf, the charge is:	\$7,040.25
The total commodity charge is:	\$19,627.25
Add the service charge:	\$80.00
The total pre-tax utility bill is:	\$19,707.25
Add state and city taxes of 5.8%:	\$1,143.02
Total bill for the month is:	\$20,850.27

Note that if the customer used no gas during the billing month, the bill would have been only the service charge of \$80.00 plus the state and city taxes for a total monthly bill of \$84.64.

TYPICAL ELECTRIC BILL CALCULATION

Fig. 2 is a sample electric rate tariff for a seasonally differentiated electric rate. In this example, assume that the customer's electricity usage in August was 200,000 kWh and the peak demand was 1655 kW. The energy rate is adjusted to account for the fuel adjustment charge (FAC) and the nuclear decommissioning surcharge. To the base rate of \$0.05890/k Wh, \$0.00123 is subtracted to account for the fuel adjustment and \$0.00074/k Wh is added to account for the surcharge. The resulting net energy charge is \$0.05841. Note that the FAC varies each month and can be either positive or negative.

The monthly charge for energy is therefore:

200,000 kWh x \$0.05841 per kWh	=	\$11,682.00
The monthly charge for demand is:		
1,655 kW x \$10.02 per kW	=	\$16,583.10
Add the service charge:		\$71.29
The resulting total pre-tax utility bill is:		\$28,336.39
Add state and city taxes of 5.8%:		\$1,643.51
The total bill for the month is:		\$29,979.90

If the customer had used no electricity during the billing month, and the highest demand in the previous 11 months

	Summer Period*	**Other Periods**
Monthly service charge	\$71.29	\$71.29
Demand charge	\$10.02/kW	\$8.53/kW
Energy charge	\$0.05890/kW	\$0.05242/kWh

Minimum monthly charge: the customer charge plus \$4.57 per kW of the highest billing demand established during the 12 months ending with the current month.

Fuel adjustment charge (FAC):	(\$0.00123 per kWh)
Nuclear decommissioning surcharge:	\$0.00074 per kWh
State tax:	4.3%
City tax:	1.5%

*Summer Period is defined as the billing months of June, July, August, and September. All other billing months are defined as "Other Periods."

Fig. 2 Sample electric rate tariff.

is assumed to be 1890 kW, the pre-tax bill would be only the service charge of $71.29 plus a minimum (demand ratchet) charge of:

$$1890\ \text{kW} \times \$4.57\ \text{per kW} = \$8637.30$$

for a total of $8708.59. Adding on the state and city tax of 5.8% results in a final bill of $9213.69 for the month. This extreme example illustrates the importance of accounting for all elements of the rate structure. Had the demand charge been ignored, the bill calculation would have been grossly underestimated.

DETERMINING THE WEIGHTED AVERAGE COST OF POWER

In the following pages, three electric rate examples are discussed. To keep the analysis manageable, certain billing factors, such as customer charges, taxes, and power factor, have been excluded. These examples, which represent typical industrial, institutional, and large commercial electric rate structures and clearly demonstrate the relationship between varying consumption load profiles and electricity costs, are based on the following three rate structure types:

- Rate 1. A seasonal time-of-use rate
- Rate 2. A conventional seasonal (CONV) rate
- Rate 3. A four-tier seasonal real-time pricing (RTP) rate.

The three rates presented here are representative of current rates in many parts of the country (between $0.05 and $0.06/kWh for baseloaded usage, inclusive of demand charges). However, they should not be used to evaluate specific technology applications. One must always use the rates charged by the local utility. It must be noted that electric rates vary dramatically across the country. In fact, neighboring utilities in the same state often have significant differences between the types of rates offered and rate levels. These differences will likely increase as the utility industry continues to undergo restructuring. While all the three rates have fairly similar costs for baseloaded usage, they have very different structures.

Rate 2 is referred to as a conventional electric rate because it has historically been the most common type of rate. It is, however, being increasingly replaced by time-of-use (TOU)-differentiated rates designed to send market price signals that shape consumer usage patterns and better reflect the cost to serve. Because the usage charge per kilowatt hour does not vary with time of use and because peak demand charges are more moderate, Rate 2 price signals do not strongly drive usage away from peak periods or attract usage in off-peak periods to the extent TOU rates do.

In the two standard (i.e., Rate 1 and 2) rates, a demand charge combines generation, transmission, and distribution system capacity charges, although each is charged separately in many rate structures. Many TOU rate structures will use varying demand charges in each rate period. In addition to peak demand charges, this TOU rate charges for excess demand in off-peak periods. The peak usage charges also include some allocation for capacity costs. But, in many rate structures, costs between peak and off-peak usage are not nearly so differentiated. In those rate structures, a larger portion of the various capacity costs are embedded in demand charges. Rates 1 and 2 have demand ratchets, which can only be set in summer months, as part of their seasonal differentiation. Many rate structures do not have ratchets and some have ratchets that can be set in any month. The RTP rate combines all capacity and commodity costs into usage charges, differentiated by four rate periods.

WEIGHTED AVERAGE COST FOR REPRESENTATIVE OPERATING LOAD PROFILES

The simple average price of electricity or gas for a given facility can be calculated by dividing the annual cost by the annual usage in billing units (e.g., kilowatt hour or 100 cubic feet). This yields an average cost per kilowatt hour or 100 cubic feet. However, average cost calculations provide limited and often misleading information about the actual incremental cost of a particular end-use or load profile. The weighted average cost for specific usage profiles may vary dramatically. In fact, it could be several times greater with one rate structure compared with another.

To demonstrate this important concept, a table reflecting the price of purchasing electricity under various usage profiles is presented for each of the three example rates. Each of these tables lists ten different power usage profiles that might be associated with usage of a certain device or, perhaps, the usage of an entire facility.

The individual profiles in Tables 1–3 show the annual usage, in kilowatt hour, for each profile, the weighted average incremental cost of a kilowatt hour, and the annual cost of consuming power under specific load profiles for a theoretical kilowatt device. Explanation of how the various profiles in Tables 1–3 are calculated and how they relate to various types of usage follows the three electric rate examples.

While different rate designs result in widely varied costs under different usage profiles, the weighted average cost for the baseloaded kilowatt usually is fairly similar for a given utility's cost structure. The rate structures and costs used in these three rate examples could all realistically be offered by one utility. The baseloaded cost per kilowatt of capacity requirement is based on

Table 1 Billing effect of 1 kW, with usage under different usage profiles operating on electric Rate 1 time-of-use (TOU)

Profile number	Period of use	Annual kilowatt hour (kWh)	Average cost ($/kWh)	Annual cost ($)	Annual load factor (LF) (%)
1	1 ratchet setting kilowatt hour per summer month	4	24.870	99	0.1
2	Baseload (BL) summer peak (no ratchet set)	700	0.137	96	8.0
3	50% LF summer peak (w/ratchet)	350	0.351	123	4.0
4	BL summer peak and shoulder (w/ratchet)	1400	0.134	187	16.0
5	6 month cooling profile (w/ratchet)	1870	0.103	193	21.3
6	12 month cooling profile	3379	0.077	260	38.6
7	50% LF 12 months peak	1043	0.169	176	11.9
8	BL 12 months all rate periods	8760	0.056	494	100.0
9	Mixed use (MU) 12 months all rate periods	4755	0.068	321	54.3
10	BL 12 months off-peak and shoulder (no peak demand)	6674	0.038	254	76.2

Notes: Load factor (LF): ratio of actual use vs. maximum potential use in all or certain rate periods; Baseload (BL): 100% LF, or the maximum use in rate period(s); Mixed use (MU): usage based on 80% peak; 60% shoulder and 40% off-peak usage; Cooling profile: demand based on 100% in 2 summer months, 85% in 2 summer months, and 60% in non-summer months. Summer usage based on 80% peak, 60% shoulder, and 30% off-peak. Non-summer usage based on 48% peak, 36% shoulder, and 18% off-peak.

continuous usage every hour of the year, with the total usage being 8760 kWh/kW of demand. This type of usage is shown for each rate example as Profile 8. This particular profile is illustrated graphically in Fig. 3. As shown, 1 full kilowatt hour is consumed in each of the 24 h in each day in each of the 12 months of the year, producing a volume of 100% of usage for one kilowatt of demand. Hence, the terms baseloaded kilowatt and 100% load factor (LF) are applied. A decrease in usage volume per kilowatt of demand corresponds to a decrease in LF. The weighted average cost per kilowatt hour for the baseloaded kilowatt is $0.0600 in the CONV rate, $0.056 in the TOU rate, and $0.056 in the RTP rate. Since the weighted average cost for baseloaded usage is close, comparison of these three rates clearly demonstrates the cost impact of rate design on various types of usage patterns.

Profile 4 in each of the rates is based on a total annual usage of only 1400 kWh for the 1 kW device. All of this usage is in the peak and shoulder rate periods during the four ratchet-setting summer months. As a result, in each of the three rates, the weighted average cost per kilowatt hour is significantly higher than the weighted average cost of the baseload (BL) usage associated with Profile 8, which also includes off-peak usage. The weighted average cost is so much higher, because it is more expensive to provide power during peak periods than off-peak periods. This is

Table 2 Billing effect of 1 kW, with usage under different usage profiles operating on electric Rate 2 conventional (CONV)

Profile number	Period of use	Annual kilowatt hour (kWh)	Average cost ($/kWh)	Annual cost ($)	Annual load factor (LF) (%)
1	1 ratchet setting kilowatt hour per summer month	4	22.850	91	0.1
2	Baseload (BL) summer peak (no ratchet set)	700	0.108	76	8.0
3	50% LF summer peak (w/ratchet)	350	0.312	109	4.0
4	BL summer peak and shoulder (w/ratchet)	1400	0.116	163	16.0
5	6 month cooling profile all rate periods (w/ratchet)	1870	0.097	181	21.3
6	12 month cooling profile all rate periods	3379	0.075	252	38.6
7	50% LF 12 months peak	1043	0.147	154	11.9
8	BL 12 months all rate periods	8760	0.060	522	100.0
9	Mixed use (MU) 12 months all rate periods	4755	0.066	315	54.3
10	BL 12 months off-peak	6674	0.048	318	76.2

Table 3 Billing effect of 1 kW, with usage under different usage profiles operating on electric Rate 3 real-time pricing (RTP)

Profile number	Period of use	Annual kilowatt hour (kWh)	Average cost ($/kWh)	Annual cost ($)	Annual load factor(LF) (%)
1	1 kWh per summer month (peak)	4	0.750	3	0.1
2	Baseload (BL) summer (peak)	700	0.201	141	8.0
3	50% LF summer (peak)	350	0.234	82	4.0
4	BL summer (peak)	1400	0.152	213	16.0
5	6 month cooling profile	1870	0.102	190	21.3
6	12 month cooling profile	3379	0.074	249	38.6
7	50% LF 12 months (peak)	1043	0.142	148	11.9
8	BL 12 months	8760	0.056	493	100.0
9	Mixed use (MU) 12 months	4755	0.066	316	54.3
10	BL 12 months (base)	6674	0.039	259	76.2

Notes: Peak: indicates that kWh are first charged to the highest rate block and then successively to lower rate blocks; Base: indicates that kWh are first charged to the lowest rate block and then successively to higher rate blocks; Cooling: profile kWh are allocated between rate blocks to correspond to the allocations used in the time-of-use (TOU) rate examples; Mixed-use: profile kWh are allocated between rate blocks to correspond to the allocations used in the TOU rate example.

reflected in the rate structures, though to varying degrees. The BL usage profile of Profile 8 blends this high-cost peak usage with low-cost off-peak usage.

The annual usage for Profile 4 is illustrated graphically in Fig. 4. Note that usage is only shown during the four summer months and during hour 6–21 of each day, which correspond to the peak and shoulder periods (6 A.M.–9 P.M.) from Monday to Friday. Hence, this figure only represents the usage during the normal five-day workweek.

In comparison to Profile 8, which shows an annual consumption volume of 8760 kWh, Profile 4 only shows a volume of 1400 kWh for the same 1 kW of peak demand. As will be shown below in the computations provided in the detailed discussion of each rate profile, the LF for Profile 4 is only 16% since only 1400 of a possible 8760 kWh are consumed over the course of the year.

In contrast, Profile 10 is based on a total annual usage of 6674 kWh, with all usage in the off-peak and shoulder rate periods. As a result, in each of the rates, the weighted average cost per kilowatt hour is even lower than the weighted average cost of the BL usage profile (Profile 8). In this case, since it is far less costly to provide electricity in the off-peak period, the result is a very low weighted average cost. The utility now has this extra capacity

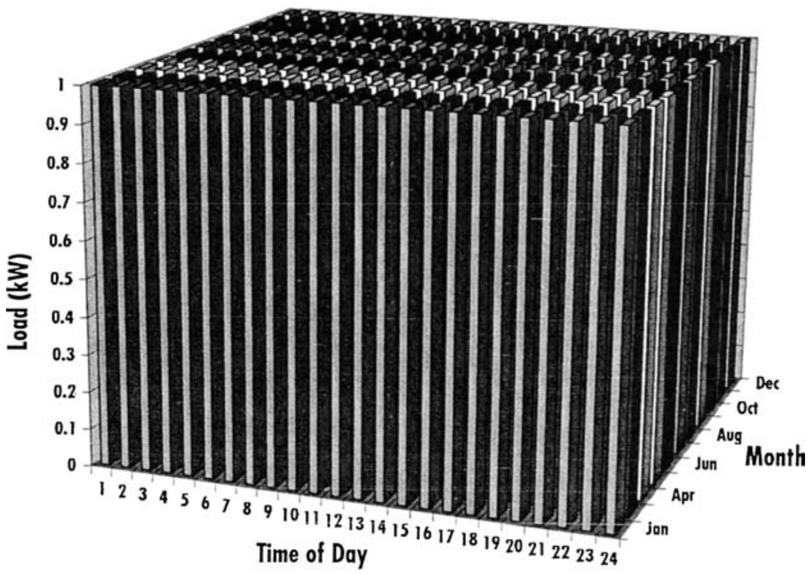


Fig. 3 Rate structure Profile 8, the baseloaded kilowatt.

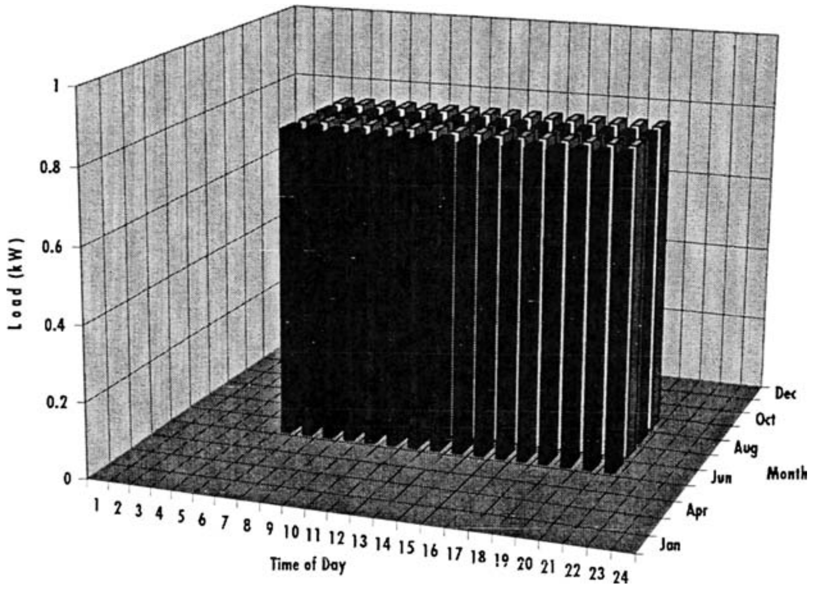


Fig. 4 Rate structure Profile 4, peak and shoulder summer usage Monday–Friday.

available, during the peak periods, that it can sell at the much higher rate to balance the sale of this low-cost usage.

The ten profiles listed in each of the Tables 1–3 were created to match the hours in each of the rate periods specific to the TOU rate structure. To allow for a reasonable basis of comparison, the same rate periods have been imposed on the rate structure used in the RTP rate. While the RTP rate has been calibrated for this purpose, it would also be appropriate to view these rate structures with respect to their own natural rate blocks.

For example, the RTP rate has a total annual base period, or lowest cost rate block of 3020 h/year, while in the TOU rate there are 4588 annual hours in the off-peak rate block. To make the RTP rate compatible with this load profile, it was assumed that the 4588 annual hours would be composed of 3020 h from the base period, with the remaining 1568 h assigned to the intermediate, or next lowest, cost rate period. The rest of the profiles for the RTP rate were calibrated to the TOU rate structure in a similar manner.

The table (i.e., Tables 1–3) accompanying each rate provides a calculation of the annual electric bill that would result from each of the ten usage profiles under each of the rate structures. Following the three rate descriptions are explanations for each of the profiles.

Electric Rate 1 (TOU)
Seasonal, TOU Electric Rate

	Summer (4 months)	Non-summer (8 months)
	Demand Charges	($/kW)
Peak	$12.00	$8.00
Shoulder Excess	$6.00	$4.00
Off-Peak Excess	$3.00	$3.00
	Energy Charges	($/kWh)
Peak	$0.068	$0.058
Shoulder	$0.058	$0.048
Off-peak	$0.032	$0.032

Fig. 5 Electric Rate 1 tariff summary.

ELECTRIC RATE 1 (TOU)

Electric Rate 1 is a seasonal TOU rate. The basic tariff is summarized in Fig. 5. Under this rate, shoulder and off-peak demand charges are assessed only for demand levels that exceed that of the peak periods. For example, if summer peak usage was 1000 kW and shoulder usage was 1500 kW, the demand charge would be:

$$(1000 \text{ kW} \times \$12.00) + (500 \text{ kW} \times \$6.00)$$

$$= \$15,000 \text{ per mo.}$$

Ratchet Adjustment

Under this rate, if one month had a peak demand of 2000 kW and the next 11 months had a peak demand of 1000 kW, each of those next 11 months' demand charges would be based on 80% of the 2000 kW figure, or 1600 kW The impact over the course of a year on the customer's bills would be an additional 600 kW in monthly billable kilowatt charge over the actual demand. Over a period of one year, the customer would pay for

6600 kW in additional demand charges. This is calculated as:

$$[(2000\ \text{kW} \times 0.80) - 1000\ \text{kW}] \times 11\ \text{months}$$
$$= 6600\ \text{kW}$$

Specific Hours of Operation

Peak:
10:00 A.M.–6:00 P.M., Monday–Friday
4 summer months at 40 h/week (700 h)
8 non-summer months at 40 h/week (1386 h)
Shoulder:
6:00 A.M.–10:00 A.M. and 6:00 P.M.–10:00 P.M., Monday–Friday
4 summer months at 40 h/week (700 h)
8 non-summer months at 40 h/week (1386 h)
Off-peak:
10:00 P.M.–6:00 A.M., Monday–Friday, all day Saturday and Sunday
4 summer months at 88 h/week (1540 h)
8 non-summer months at 88 h/week (3048 h)

ELECTRIC RATE 2 (CONV)

Electric Rate 2 is a conventional seasonally differentiated Commercial, Industrial and Institutional (CI&I) rate. The basic tariff is summarized in Fig. 6. Specific hours of operation are:

Summer–4 months (17.50 weeks totaling 2940 h)
Non-summer–8 months (34.64 weeks totaling 5820 h).

ELECTRIC RATE 3 (RTP)

Electric Rate 3 is a simplified real-time-pricing rate. There are several ways in which this developing rate is offered by electric utilities to customers. In one common approach, the pricing is established per rate block based on an analysis of the utility's costs associated with the dispatch of various generation stations in the stack. Each day, the utility informs the customer of which hours will be applied to each rate block.

Electric Rate 2 (TOU)
Seasonal, TOU Electric Rate

	Summer (4 months)	Non-summer (8 months)
Demand Charges ($/kW)	$10.00	$8.00
Energy Charges ($/kWh)	$0.051	$0.046

Fig. 6 Electric Rate 2 tariff summary.

Rate Block	Unit Cost ($/kWh)	Summer Hours	Non-summer Hours	Total Hours
1. Base	$0.025	700	2,320	3,020
2. Intermediate	$0.038	1,200	3,200	4,400
3. Peak	$0.170	1,000	300	1,300
4. Power pool peak	$0.740	40	0	40
Total		2,940	5,820	8,760

Fig. 7 Real-time pricing (RTP) rate pricing blocks.

In this case (Fig. 7), four blocks have been assigned: base, intermediate, peak, and power pool peak. The power pool peak refers to costs incurred as a result of the utility requiring peak power from the pool. Since each hour of the year is assigned to a rate block based on actual (real time) dynamic conditions, there is no established schedule. A good approximation can be made based on experience, however. For the purpose of demonstrating the workings of this rate, hours have been assigned for winter (36.64 weeks) and summer (17.50 weeks) to each of the four rate blocks.

Fig. 8 shows the maximum cost in each rate block for a baseloaded kilowatt Fig. 9 shows the annual hours of operating and cost per kilowatt hour for various combinations of rate blocks.

ELECTRIC RATE COMPARISONS AND CONCLUSIONS

The period during which power is used affects operating costs as much as, or more than, the amount of power used. This concept becomes critical in energy use planning as price differentiation by time of use increases.

While accountants may look at electric operating costs in terms of the average cost per kilowatt hour, energy use planners must look at the incremental costs of individual

Rate Block	Summer	Non-Summer	Annual
1. Base	$17.50	$58.00	$75.50
2. Intermediate	$45.60	$121.60	$167.20
3. Peak	$170.00	$51.00	$221.00
4. Power pool peak	$29.60	$0.00	$29.60
Total	$262.70	$230.60	$493.30

Fig. 8 Maximum cost per rate block.

Annual and Average Costs			
			Weighted
Rate Block	Hours	Ann. Cost	Avg. Cost
		($)	($ kWh)
3+4 Summer	1,040	$199.60	$0.1n'
2+3+4 Summer	2,240	$245.20	$0.109
1 +2+3+4 Summer	2,940	$262.70	$0.089
1 +2 Annual	7,420	$242.70	$0.033
1 +2+3 Annual	8,720	$463.70	$0.053
3+4 Annual	1,340	$250.60	$0.187
2+3+4 Annual	5,740	$417.80	$0.073
1+2+3+4 Annual	8,760	$493.30	$0.056

Fig. 9 Operating hours and cost per rate block combinations.

end uses and various consumption profiles to understand price impact. Energy planners audit facilities to develop incremental cost-usage profiles associated with individual equipment, systems, and activities. These audits are done in much the same manner as the ten profiles presented in the preceding pages.

Planners look at seasonal end uses, such as cooling, and understand that the relevant weighted average cost per kilowatt hour may be several times greater than the facility's overall average cost per kilowatt hour, especially if a ratchet adjustment is in effect. They look at baseloaded operations and consumption blocks and understand that the costs may be lower than the facility's average cost. They look at identical devices, in a multiple-unit system, that run the same amount of hours per year, and they understand that if they operate with different load profiles, their operating cost may be dramatically different.

Evolving RTP electric rate structures may extend this discrete differentiation to every hour of the year, or perhaps even every minute. An important benefit of RTP rates is that one peak hour of extraordinary usage might not have the dramatic cost impact that a rate with a high demand charge and ratchet adjustment would have. This type of rate flexibility is well suited for electricity purchase strategies that involve a mix of on-site power generation or purchase of non-utility-generated power along with the purchase of utility provided power. In the event of retail purchase of non-utility-generated electricity, traditional, TOU or RTP type rate structures may be applied to transmission and distribution services, while some type of RTP structure would be applied to the usage for the purpose of commodity transaction.

With this understanding, energy planners develop strategies to minimize operating costs and optimize productivity. Efficiency improvements and alternative energy source options are considered with respect to these incremental costs. Self-generation and electricity displacement strategies should be evaluated in the same

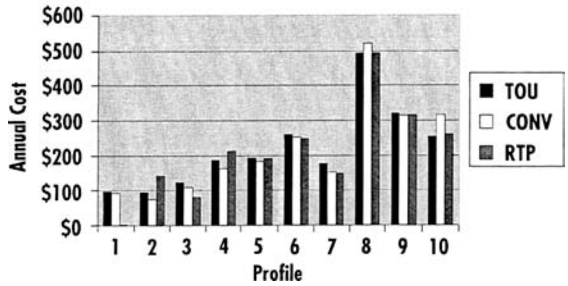


Fig. 10 Annual operating cost for each load profile.

manner. Electric cost savings opportunities should be considered on the basis of incremental cost per kilowatt hour, as well as the total usage. Elimination of 1 kW of low-LF usage should produce larger cost savings per kilowatt hour than the elimination of 1 kW of high-LF usage, even though it may not produce greater aggregate energy savings.

Following are explanations for each of the profiles and a discussion of the type of equipment usage or facility characteristics that would result in each profile. Also included are examples of how LF, annual cost, and weighted average cost were calculated.

DETERMINING THE WEIGHTED AVERAGE COST FOR VARIOUS LOAD PROFILES

For individual equipment or an entire facility operating with anyone of the load profiles presented in Tables 1–3, the total annual usage and cost are based on the total input power (kilowatt) of the equipment (or the connected load of the facility) times the usage and cost of 1 kW as presented in each table entry. In all cases, it is assumed that the facility has only one billing meter that measures consumption and demand of all connected loads.

The annual and weighted average costs per kilowatt hour for the ten sample load profiles under each of the example utility rates are summarized in Figs. 10 and 11. Note that because of the extremely low LF for Profile 1,

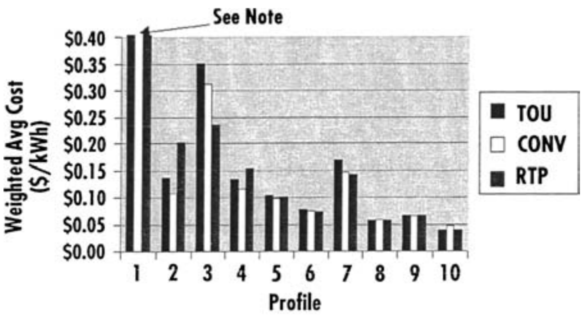


Fig. 11 Weighted average cost ($/kWh) for each load profile.

TOU and CONV average costs shown in Fig. 11 are off the scale (>$20/kWh).

EXPLANATION OF TEN SAMPLE LOAD PROFILES

Profile 1 is based on a device rated at 1 kW, operated only one hour during each summer month June, July, August, and September) when the entire facility is already operating at its highest peak demand level. This added load increases the peak demand level for the month by 1 kW Therefore, there is a peak demand charge for that 1 kW in each of the 4 summer months. It also adds 1 kW to any applicable peak demand ratchet level. Under Rate 1, for instance, the effect of this 1 kW is a monthly demand charge based on 0.8 kW in each of those 8 non-summer months.

Thus, the 4 kWh produce a total annual billable demand charge based on 10.4 kW. In addition, there is a usage charge for the 4 kWh totaling $0.27. Based on Electric Rate 1, the annual cost for using the 4 kWh is:

$$(1\ \text{kW} \times \$12/\text{kW} \times 4) + (0.8\ \text{kW} \times \$8/\text{kW} \times 8) + (4\ \text{kWh} \times \$0.068/\text{kWh}) = \$99.47$$

The weighted average cost per kilowatt hour is:

$$\$99.47/4\ \text{kWh} = \$24.87/\text{kWh}$$

If that same 1 kW device is operated for 4 h in the off-peak rate period, then no demand charges apply and the consumption charges are lower. The total annual cost of the 4 kWh is only $0.13, and the weighted average cost of 1 kWh is $0.032/kWh. While the first profile is an extreme example, it demonstrates how significant the impact of demand charges and ratchet adjustments can be in a given rate structure.

A large chiller used for space cooling, for example, may operate at peak capacity only a few hours during the entire year. In many cases, peak cooling demand coincides with the facility's maximum electric use period and establishes not only a demand peak for the month, but also an increased ratchet demand level for the year. If, at the hour of maximum monthly electric usage (the peak demand hour), the chiller consumes 1 extra kilowatt hour to satisfy load, that 1 kWh will increase the billing peak demand by 1 kW.

Taking the most extreme case of Profile 1 in Rate 1, the use of only 4 kWh at the peak demand period of the peak demand month would cost $99. This could actually be the case with a resting lab or a university that holds commencement in the summer and experiences an extraordinary peak load on only one day per year. If the facility is able to somehow shed these 4 kWh at the same time that the chiller or testing equipment consumes these ratchet-setting kilowatt hour, it would save $99.

At first glance, the idea of 4 kWh costing $99 might seem absurd. However, an understanding of how electric rate structures operate shows that the incremental cost per kilowatt hour can actually vary by several thousand percent. In fact, with a typical demand window of 15 min, this equipment need only set a peak for 15 min, consuming only 0.25 kWh to cause the facility to endure the full 10.4 kW of annual demand charge. Under the RTP rate, no such dramatic costs would be incurred. Since there is no demand charge impact, the costs only consist of the peak cost per kilowatt hour times the hours of use. In this example, the most expensive kilowatt hour of the year would cost $0.74 and the 4 kWh would, therefore, cost $2.96. However, it is important to note that in a real-time market, the cost for these 4 kWh could be quite a bit higher, conceivably as high as $99, though perhaps not likely.

Profile 2 is based on a device rated at 1 kW and operated as a BL in all 4 summer month peak hours. Under Rate 1 (the TOU example consisting of 700 h/year based on a 40-h/week rate period extending for 17.5 weeks), the annual cost for using those 700 kWh is:

$$(1\ \text{kW} \times \$12/\text{kW} \times 4) + (700\ \text{kWh} \times \$0.068/\text{kWh}) = \$96$$

The weighted average cost per kilowatt hour is:

$$\$96/700\ \text{kWh} = \$0.137/\text{kWh}$$

The annual LF is:

$$700\ \text{h}/8760\ \text{h} = 8\%$$

In this scenario, demand charges total $48, slightly more than half of the annual cost. It is assumed in this case that peak demand is sufficiently high in the winter months so that there is no ratchet in effect. Based on Rate 1, if, for example, an electric motor with an input power rating of 100 kW operates at full load in each of the 700 h in this profile (represented in Table 1, Profile 2), the annual cost would be $9600. This can also be calculated by multiplying the input power (100 kW) by the total full-load hours of operation (700 h) by the weighted average cost per kilowatt hour of $0.137.

Under the RTP rate, assuming that the 700 h corresponding to the TOU rate summer peak would be composed of the full 40 h of the power pool peak block and 660 h of the peak block, the total cost would be $141 and the weighted average cost would be $0.20/kWh. Based on this example, operating the same 100 kW electric motor during the same 700 h would result in an annual cost of $14,350.

In Profile 3 the 1 kW device operates with a 50% LF during the 40 h/week peak rate period over the 4 summer months rather than at a 100% LF. This results in a usage of

350 kWh over the 700 total hours in this rate period and a total annual LF of 4%, rather than 8% with Profile 2. Usage over the 17.5 week summer period and annual LF, respectively, are calculated as follows:

$$(1\text{ kW} \times 17.5\text{ weeks}) \times (40\text{ h/week}) \times (0.5\text{ LF}) = 350\text{ kWh}$$

$$\frac{700\text{ h} \times 0.5\text{ LF}}{8760\text{ h}} = 4\%$$

Using the TOU rate, the total annual cost is reduced compared with Profile 2 due to reduced usage. Therefore, demand charges as a percent of the total cost increase. In Profile 3 (with ratchet adjustment), demand charges represent 84% of the total cost. The impact of reduced usage with constant demand charges is an increased weighted average cost per kilowatt hour. The weighted average cost per kilowatt hour for Profile 3 increases to $0.351.

For this profile with only summer peak usage, the traditional non-time-differentiated rate is less costly than the TOU rate, and the RTP rate is the lowest, with a weighted average cost of $0.234. The reason is that with such a low LF load of 4%, the impact of demand charges, inclusive of ratchets, drives up costs dramatically for the other rate structures.

Profiles 2 and 3 are realistic examples of the cost of cooling equipment operation during peak periods. In many cases, these operating profiles are only a portion of a cooling unit's total energy usage. In other cases, they may represent the total operation. In facilities with multiple cooling units, one unit is often predominantly used as a peaking unit. Peak cooling loads often correspond to the TOU peak electric rate period (10:00 A.M.–6:00 P.M., Monday–Friday) due to ambient temperature profiles, increased productivity, and increased internal gains from people and equipment. In single shift C&I operations, the peak period coincides with most of the operating hours of the facility. In those cases, profiles such as Profiles 2 and 3 may also be representative.

Electric usage with profiles of the type listed in Profiles 1–3 are often targeted for elimination or reduction by various load shedding or alternative energy source technologies. Peak-shaving generators and fuel- or steam-powered cooling are two commonly applied technologies for eliminating these blocks of electric usage.

Profile 4 is similar to Profiles 2 and 3, in that it reflects the higher cost of power resulting from seasonal differentiation and significant demand charges. With 100% usage in Summer Peak and Shoulder periods, this profile has greater usage, with a LF of 16%, as compared with 4% in Profile 3. Under the TOU rate, this produces a weighted average cost per kilowatt hour of $0.134, as compared to $0.351 for the lower LF usage of Profile 3.

In this profile, demand charges represent a lower percentage of the total cost than in Profile 3 because each unit of demand is spread over a greater usage base. However, demand charges still represent a significant portion of the total cost. Profile 4 also includes the effect of demand ratchets. Extending usage to the shoulder periods in this profile partially integrates lower-cost power into the profiles and results in increased LF and decreased weighted average cost.

Under the RTP rate, assuming the 1400 h in Profile 4 would include 40 power pool peak hours, 1000 peak hours, and the balance of 360 h intermediate, the annual cost and weighted average cost, respectively, are:

$$(40\text{ kW} \times \$0.74) + (1000\text{ kWh} \times \$0.17) + (360\text{ kWh} \times \$0.038) = \$213.28$$

$$\frac{\$213.28}{1400\text{ kWh}} = \$0.152/\text{kWh}$$

Profiles 5 and 6 refer to mixed use cooling season profiles. While Profiles 2–4 are all based on a 1 kW device running either all of the time or with a 50% LF in a given rate period, Profiles 5 and 6 are based on a 1 kW cooling device running with a load that varies between each month and rate period. These profiles were designed to represent typical space cooling loads served by an electric vapor compression system.

During the 4 summer months, it is assumed that the equipment operates with a LF of 80% peak, 60% shoulder, and 30% off-peak, based on the TOU rate structures. In July and August, it is assumed that the full 1 kW peak demand is set. In June and September, it is assumed that the peak demand impact of the 1 kW equipment is 0.85 kW In the non-summer months, it is assumed that the equipment operates with a LF of 48% peak, 36% shoulder, and 9% off-peak and the demand impact is 0.60 kW in each month. These profiles were then calibrated and applied to the CONV and RTP rate. The difference between the two profiles is that Profile 5 is based on 6 months of operation and profile 6 is based on year-round operation. Notice that profile 6 has a far lower weighted average cost per kilowatt hour than Profile 5 as more lower cost non-summer usage is blended in and there is no ratchet impact.

Profile 7 has peak usage every week of the year and demand charges in every month. This type of profile might be targeted for elimination with peak shaving power generation or replacement of baseloaded electric-driven equipment with fuel- or steam-driven equipment operated in the peak periods.

Profile 7 is representative of a load profile that might result from electric-driven equipment operated with a 50% LF during the peak period only. Under the TOU rate, this corresponds to 20 h of operation per week. There is a significant increase in weighted average cost for Profile 7 for all rate types. This is due to the fact that demand charges remain the same, but are spread over only half the usage as the LF is only 11.9%. Under the RTP rate, the weighted average cost is relatively high because as LF is decreased, a greater percentage of the usage is assumed to fall in the highest cost rate block.

Profile 8 (BL 12 months, all rate periods) represents the baseloaded kilowatt, or 1 kW baseloaded every hour of the year. The annual usage of 8760 kWh has an annual LF of 100%.

With this profile, the fixed investment in electric generation and distribution capacity is spread over the maximum possible annual usage. As shown in Profiles 1–7, different rate structures result in widely varied costs under different usage profiles. However, the weighted average costs for the baseloaded kilowatt usually are fairly similar for different rates under a given utility's cost structure. The rate structures and costs used in the TOU, CONV and RTP rates could realistically be offered by one utility.

While with the profiles with lower LF, peak usage produced much higher costs for the RTP rate, compared with the CONV rate, the higher LF off-peak usage produced much lower costs with the RTP rate. These opposing trends are roughly canceled out with continuous BL usage, though traditional rate structures, such as Rate 2, commonly produce slightly higher baseloaded kilowatt hour costs. Thus, a three-shift facility with a very high LF would choose the TOU or RTP rate structures.

The annually baseloaded kilowatt is the load profile most often targeted for prime mover-driven power generation and mechanical service applications that employ heat recovery (cogeneration cycles). The weighted average cost per kilowatt hour of the baseloaded kilowatt is often the benchmark for determining the feasibility of such applications. While the cost per kilowatt hour is lower than most of the other load profile entries, the total annual cost is the greatest.

Profile 9 represents a profile that might result from annual operation of individual electric motors or other process equipment. It might also result from the combined operation of multiple equipment in a facility. Facilities rarely have absolutely flat loads. Therefore, the BL, or 100% LF load profile, is not necessarily representative of the weighted average cost of power. Some type of mixed use profile, such as Profile 9, is generally more representative of the weighted average cost. This is an aggregate of varying individual components, such as lights and motors, each with a different load profile.

Sometimes, equipment does operate with a LF of 100%, either in a specific rate period or in all rate periods. More often, equipment operates under varying load or under full load for intermittent periods. Profile 9 is an example of such operation. The LF of this profile is about half that of Profile 8. The total annual cost is lower due to significantly lower use, but is greater than half the cost of Profile 8, because the weighted average cost per kilowatt hour is greater. This is due to the greater relative impact of demand charges (the dollar value of which remains the same as in Profile 8).

Profile 10 represents extensive off-peak non-demand setting usage. It has no demand charge at all. Notice that with no demand charges, the weighted average costs are lower for the TOU rate than the CONV rate. This is a result of the price signals of the TOU differentiated rates, which greatly emphasize usage in off-peak and shoulder periods—which the traditional rate does not do. Also note that only a lower cost excess demand charge can be assessed to these usage profiles in the TOU rate example, while a full demand peak can be set in the CONV rate. This would increase costs still further. Under the RTP rate, these average weighted costs are produced by blending the lowest cost rate block with the successively higher cost rate blocks.

As Profile 10 shows, TOU rates offer relatively low-cost power for about three quarters of the hours of the year. When an excess demand charge is assessed to shoulder usage, or if standard shoulder-period demand charges are in effect, the costs will be slightly greater, although still significantly lower than during the costly peak period.

The RTP rate offers relatively low-cost power for about 85% of the annual hours. The weighted average annual cost for the 3020 h of base block usage and 4400 h of intermediate block usage is only $0.033/kWh. This is balanced by a much higher unit cost in the peak and power pool peak blocks, which comprise about 15% of the total annual hours, but slightly more than half the total annual cost. The weighted average cost is only a usage cost, but it also reflects imbedded fixed costs, a large portion of which are included in the demand charges associated with the other two rates. Table 4 provides a comparison of Profiles 7, 8, and 10 for the CONV (2) and RTP (3) rates.

Notice that the sum of total costs for Profiles 7 and 10 approximately equal Profile 8, the annually baseloaded kilowatt The total annual costs of Profiles 7 and 10 are somewhat dose, compared with the stark contrast in usage. The weighted average cost per kilowatt hour for Profile 7 is about three times that of Profile 10.

MATCHING ENERGY TECHNOLOGY ALTERNATIVES WITH ELECTRIC RATES AND USAGE PROFILES

Following is a brief discussion of the fuel- and steam-powered technology applications that should be considered for use in eliminating electricity purchases associated with the 10 representative load profiles presented in Tables 1–3.

Table 4 Comparison of Profiles 7, 8, and 10 for conventional (CONV) and real-time pricing (RTP) rates

Rate number	Profile number	Usage (kWh)	Annual load factor (LF) (%)	Average cost per kWh ($)	Total cost ($)
2	7	1043	12	0.147	154
2	10	6674	76	0.048	318
2	8	8760	100	0.060	522
3	7	1043	12	0.142	148
3	10	6674	76	0.039	259
3	8	8760	100	0.056	493

Electric Peak Shaving Generation Applications

The primary focus here is on peak demand and usage charge savings. Thus, the emphasis is on eliminating costly peak demand charges resulting from poor LF or, in the case of RTP rates, eliminating the usage in the highest cost rate blocks. As shown above in the RTP rate, the total annual cost is about the same for a load with a 15% LF occurring in the highest cost rate blocks as a load with an 85% LF occurring in the lowest cost rate blocks. Since peak shaving applications will have relatively low annual hours of operation, low capital cost is emphasized more heavily than optimum thermal efficiency and simple energy costs, and heat recovery is not commonly used. Representative load profiles that might be targeted for peak shaving generation include Profiles 1–3, and 7. Other potential technology applications include load shedding control, battery storage and thermal energy storage (TES).

Electric Cogeneration Applications

The primary focus of power generation applications that employ heat recovery is on overall system thermal fuel efficiency and durability, since equipment run times may range from several thousand hours to continuous operation. The object is to minimize the use of purchased electricity while simultaneously eliminating other internal fuel usage via heat recovery. The 8760 h load profile associated with Profile 8 is ideal for cogeneration in most cases. Other high LF load profiles, such as Profile 9, may also be targeted for elimination with on-site application of cogeneration technologies. In some cases, notably with highly stratified rate structures, it may not be economical to generate power on site in periods when the lowest cost power is available from the utility. In such cases, load profiles with somewhat less than 100% LF may be appropriately targeted for elimination. Other potential technology applications include combined cycles and steam injection cycles, along with a host of other standard BL energy conservation measures.

Single- Unit, Year-Round Mechanical Drive Applications

The primary focus here is on satisfying end use requirements with a single unit that has the lowest life cycle cost. Single-unit systems may operate several thousand hours or more annually. In some cases, prime mover-driven systems using heat recovery may be less costly to operate than electric-driven systems in all use periods. In other cases, prime mover-driven systems may be more costly in some periods (i.e., off-peak), but less costly to operate on the average (mixed use) due to significant savings in peak and shoulder periods. System efficiency and durability are particularly emphasized for applications with significant run-time requirement. Heat recovery will be increasingly cost-effective with increased hours of operation. In single or two-shift operations, the unit may operate only in costly peak periods or in peak and shoulder rate periods. Representative load profiles that might be targeted for elimination with single-unit prime-mover mechanical drive systems include Profiles 8 and 9. Other potential technology applications include building and process automation systems and variable volume distribution systems (i.e., air, water, steam, etc.).

Multiple-Unit Mechanical Drive Mixed (Hybrid) System Applications

The primary focus here is overall system optimization with use of electric units in off-peak periods and non-electric units during peak and shoulder periods, or a variation with some baseloading of equipment. Equipment may operate anywhere from several hundred hours to several thousand hours annually. Thermal fuel efficiency (and heat recovery potential) of non-baseloaded individual units may be sacrificed for lower capital costs.

A cogeneration cycle unit may be baseloaded and electric units used for the remaining off-peak load and non-electric units used for the remaining peak and shoulder loads. Alternatively, an electric unit may be baseloaded and non-electric units used for peaking duty only. In one- or two-shift, five-day operations, a single unit might experience a similar operating profile as would a peaking unit in a three-shift, seven-day operation. Profile 7 might be targeted for elimination with use of a prime mover-driven mechanical drive system. Other potential technology applications include building and process automation, peak shaving, load shedding control and variable volume distribution systems (i.e., air, water, steam, etc.).

Single-Unit Seasonal Cooling Applications

The primary focus here is satisfying cooling requirements with a single unit with the lowest life cycle cost. Equipment may operate anywhere from several hundred hours to a few thousand hours annually. Heat recovery is a viable option, but has less impact than in year-round, single-unit applications, because operating hours are typically lower. Heat recovery becomes more important with higher hours of operation and fuel costs. In single- or two-shift operations, the unit may operate predominantly in costly peak periods or peak and shoulder rate periods. Representative load profiles that might be targeted for elimination with use of a non-electric-driven cooling systems include Profiles 2–6. Other potential technology applications include peak shaving, load shedding control, battery storage and TES.

Multiple-Unit Cooling (Hybrid) Applications

The primary focus here is to minimize system operating costs with use of electric units in off-peak periods and non-electric units in peak and shoulder periods, or a variation with some baseloading of equipment. Electric peak demand charges are a critical consideration. Equipment may operate anywhere from two hundred to several thousand hours annually. Thermal efficiency (and heat recovery potential) on non-baseloaded individual units may be sacrificed for lower capital costs. This is similar to strategies for year-round, multiple-unit mechanical drive systems. However, optimization strategies will differ somewhat due to the greater electric unit costs with seasonal pricing and ratchet potential. In one- or two-shift, five-day operations, a single unit might experience a similar operating profile as would a peaking unit in a three-shift, seven-day operation. Representative load profiles that might be targeted for elimination with use of non-electric-driven cooling equipment as part of a mixed system application include Profiles 2–4. Other potential technology applications include peak shaving, load shedding control, battery storage and TES.

CONCLUSION

As shown in this article, the weighted average cost of a kilowatt hour is highly variable. Different technology applications must be matched with different cost scenarios and electric load profiles. The aggregate energy cost for any particular application is the real determinant of operating cost savings potential. While savings of $0.30 or $0.60/kWh are attractive targets, savings must accrue over enough hours to provide sufficient payback on the investment in alternative energy equipment. Baseloaded cogeneration applications, for example, emphasize overall thermal fuel efficiency, effective heat recovery, and durability. These systems will save less per hour of operation than other technology applications in many cases, but will accrue savings over a greater number of operating hours. Conversely, peak shaving applications may run fewer hours, but target the most expensive electricity usage. In these applications, thermal fuel efficiency is generally less of a concern than low capital costs, because the key is elimination of low LF high-cost power purchases.

BIBLIOGRAPHY

1. Capehart, B.L.; Wayne, C.T.; William, J.K. Understanding energy bills. *Guide to Energy Management* 5th Ed.; The Fairmont Press, Inc.: Lilburn, GA, 2006; [Chapter 3].
2. Mahoney, J.F.; Barney, L.C. Strategies for electric utility bill reduction when alternative rate structures apply. Eng. Economist **1994**, *Spring*.
3. Petchers, N. Utility rate structures, In *Combined Heating, Cooling and Power Handbook*; Marcel Dekker, Inc.: New York, 2003; [Chapter 21]. Petchers, N. Utility bill analysis; In *Combined Heating, Cooling and Power Handbook*; Marcel Dekker, Inc.: New York, 2003; [Chapter 21].
4. Turner, W.C. Electric and gas utility rates for commercial and industrial customers. *Energy Management Handbook* 5th Ed.; The Fairmont Press, Inc.: Lilburn, GA, 2005; [Chapter 18].

Utilities and Energy Suppliers: Business Partnership Management☆

S. Kay Tuttle
Process Heating, Boilers and Foodservices, Duke Energy, Charlotte, North Carolina, U.S.A.

Abstract

This paper explores the role of the utility and the industrial or commercial facility engineer in developing a business partnership especially as it relates to the quality of electric power. The industrial or commercial facility engineer needs a thorough knowledge of the electrical power quality needs of equipment and processes critical to the business's financial success. Information relating to the operation of the electric utility system is essential to understanding what the utility can reasonably deliver and determining the best solution for bridging any gaps between process requirements and available power quality. The utility needs a thorough understanding of the equipment and processes of businesses they serve as well as a thorough knowledge of the power quality requirements of that equipment. Also, the utility should have available the best options for each process to resolve any gaps between the quality of power supplied and that required by the process. The utility has an obligation to its business partners to explain the operations of the utility system and to make available solutions to power quality problems or needs. Some utilities now offer to install power quality monitors, diagnose power quality problems, perform wiring and grounding assessments, and procure and install mitigation equipment within commercial and industrial facilities. When an opportunity arises for the facility engineer and the utility power quality engineer to work together to prevent or resolve a power quality issue, each should be prepared with the knowledge, information and resolve to do so.

INTRODUCTION

It is imperative for utilities and their customers to form a partnership in order for both to optimize their businesses. Electricity and productivity in industry are inherently linked. For a utility to remain profitable, it must have productive customers, and for a customer to be productive, he or she must have reliable, high quality power. Partnership implies trust. The utility must trust customers and disclose the reliability and quality of power on the electrical system, and the customers must trust the utility and inform the utility engineers regarding their processes. Working together both can meet their business objectives.

SHARING INFORMATION WITH CUSTOMERS ABOUT THE UTILITY SYSTEM

Before the utility can share information regarding typical system performance, the utility must first monitor and analyze the system. In an effort to characterize the utility distribution system, Duke Power and other electric utilities funded a study to determine what quality of service a distribution served customer might expect. This study called Distribution Power Quality (DPQ) was conducted by Electric Power Research Institute (EPRI) and focused on voltage sags, transients, and harmonics. The results for voltage sags are presented in Fig. 1. This information is extremely valuable to customers when specifying new equipment or selecting the optimum power quality mitigation devices. For example, the data indicates that the great majority of voltage sags are 6–10 cycles in duration and from 60% to 90% remaining voltage. A customer can use this information when ordering new equipment by requesting equipment that can ride through these voltage sags. If the equipment manufacturer does not have the capability to test the ability of the new equipment to ride through voltage sags, the utility can do so. Or, if mitigation equipment is needed, this information is valuable when choosing which mitigation equipment to apply.

UNDERSTANDING CUSTOMER PROCESSES

A utility gains valuable information by analyzing its system, but that is only part of the information needed to make businesses more productive. A detailed understanding of customer processes and what causes these

☆This entry originally appeared as "Building Business Partnerships with Your Electric Utility: A Power Quality Perspective" in *Energy Engineering,* Vol. 101, No. 3. Reprinted with permission from AEE/Fairmont Press.

Keywords: Power quality; Power disturbances; Sags; Utility partnerships; Equipment sensitivity.

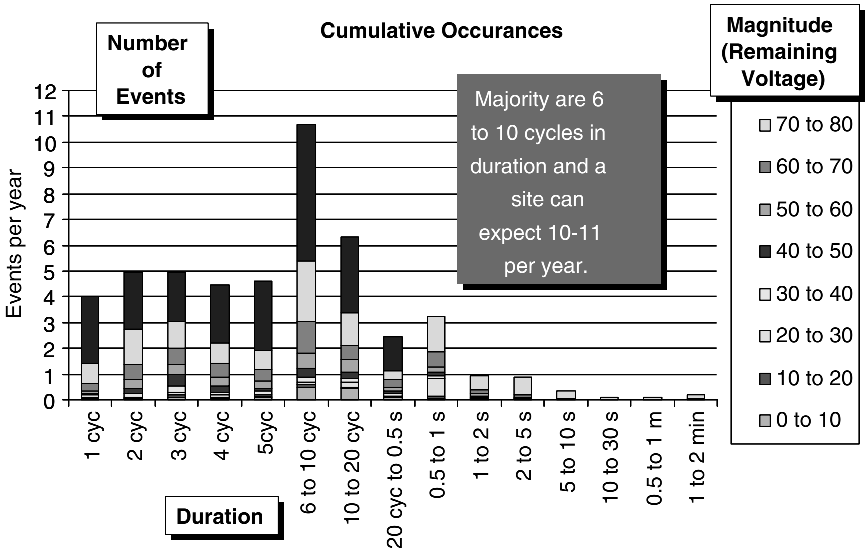


Fig. 1 Voltage sag data from the distribution power quality (DPQ) study.

processes to unexpectedly shutdown is also needed. Voltage sags are a common cause of process interruptions for many industrial customers. Voltage sags cause equipment and therefore, entire processes to shut down, impacting productivity as well as product quality. In an effort to resolve many of the problems caused by voltage sags, utilities provided funding to EPRI to perform an extensive testing of processes prone to shutdowns and to determine effective solutions. The information is in a web-based tool called the Industrial Design Guide (IDG).

EPRI and the participating utilities prioritized industries for study each year. To date, plastics and polymers processing, metals fabrication, semiconductor fabrication, pharmaceuticals, printing and publishing, and healthcare and hospitals have been addressed. The plastics and polymers processing industry was first completed. Processes within this industry were identified early and included pipe extrusion, plastic jacket extrusion, thermoforming, blown film—bubble, and injection molding. Each of these processes was tested in order to find the weak links within the process. One objective was to identify the most economical and effective power quality solutions and to do so required, identifying what components within the process caused the process to shut down. A process diagram of blown film is shown in Fig. 2. The equipment shown in dark gray is possible weak links and can be viewed in even more detail within the IDG.

A tool that enables component level testing was developed several years ago. This tool is connected to and powers the process being tested. The process is then subjected to controlled voltage sags of different magnitudes and durations, originating at various points on the voltage sinewave while the response of process components is being monitored. The result is detailed information about the ride through capabilities of various components during voltage sags. A sample of the results from such a test is shown in Fig. 3. This figure shows that for a sag to 87% nominal voltage for one cycle or longer, the photo eye, the most sensitive component in the process, will drop out causing the process to shut down. The next most sensitive component in this example is the double-pole, double-throw, DPDT, relay. A voltage sag to 75% of nominal voltage lasting two cycles or longer will cause this relay to dropout.

Once the process is characterized to the component level, solutions for each of the components can be identified as well as the cost effectiveness of each solution. With this information, the customer can make an informed business decision about the acceptable number of shutdowns per year and the cost he or she is willing to pay to achieve this level of ride-through.

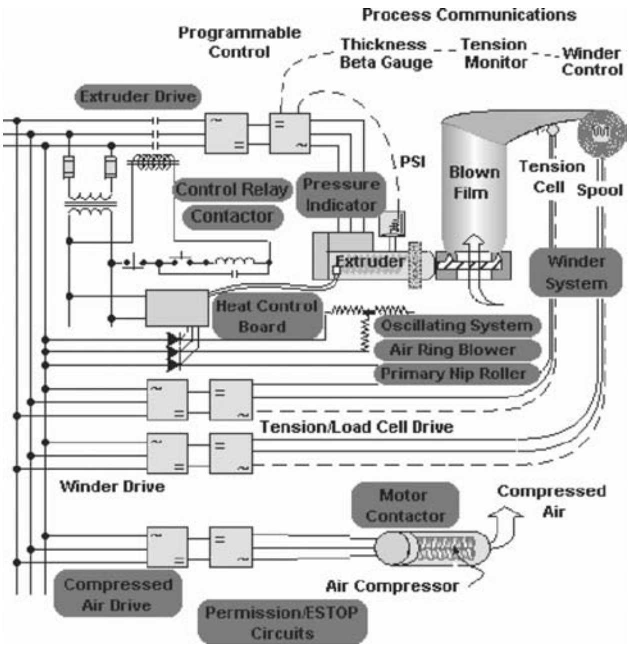


Fig. 2 Diagram of blown film process and possible weak links.

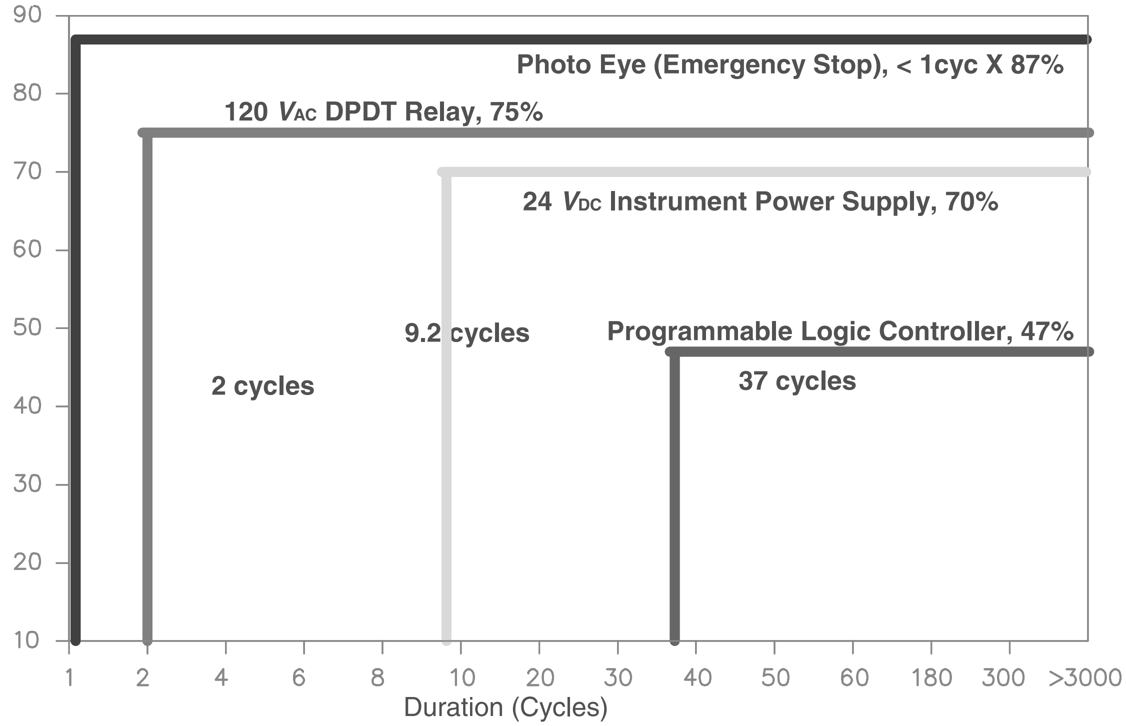


Fig. 3 Process component sensitivity.

Fig. 4 shows contour lines generated from the DPQ Study. For example, a distribution served customer with equipment sensitive to voltage sags in the highlighted area can reasonably expect between 10 and 15 process interruptions a year.

Fig. 5 overlays the process component sensitivity found from sag testing with the distribution system characterization curves from the DPQ Study. By examining the contour lines that the equipment sensitivity plot is within, a customer can predict how often a particular component

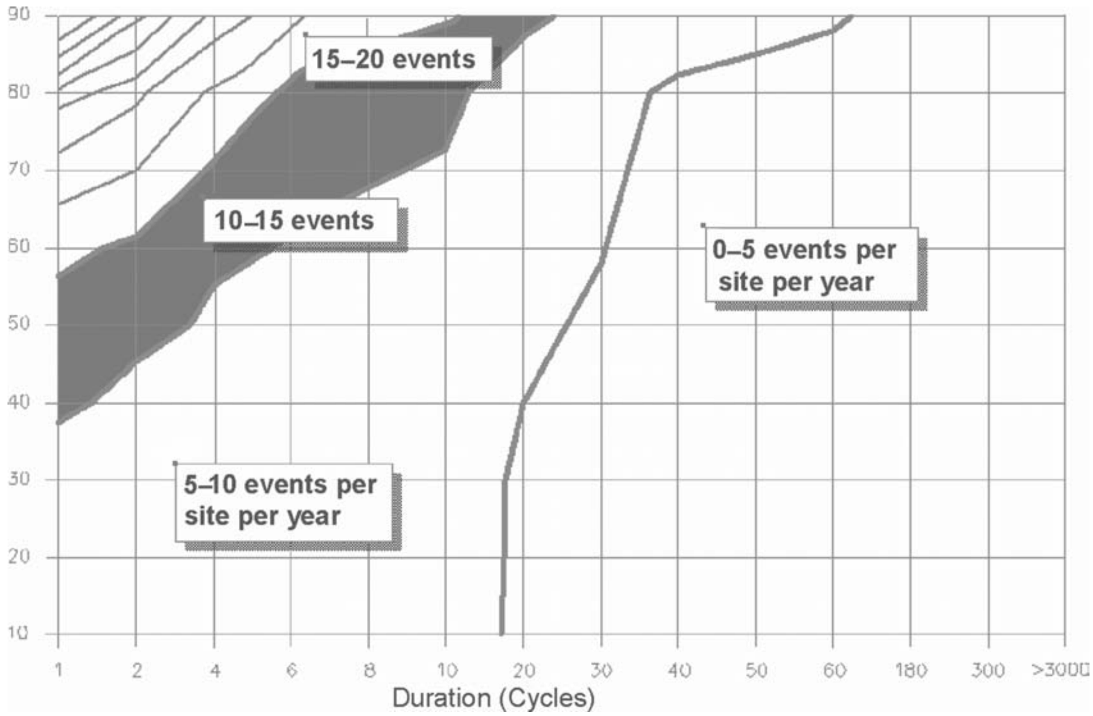


Fig. 4 DPQ study distribution system characterization curves.

Util—Wall

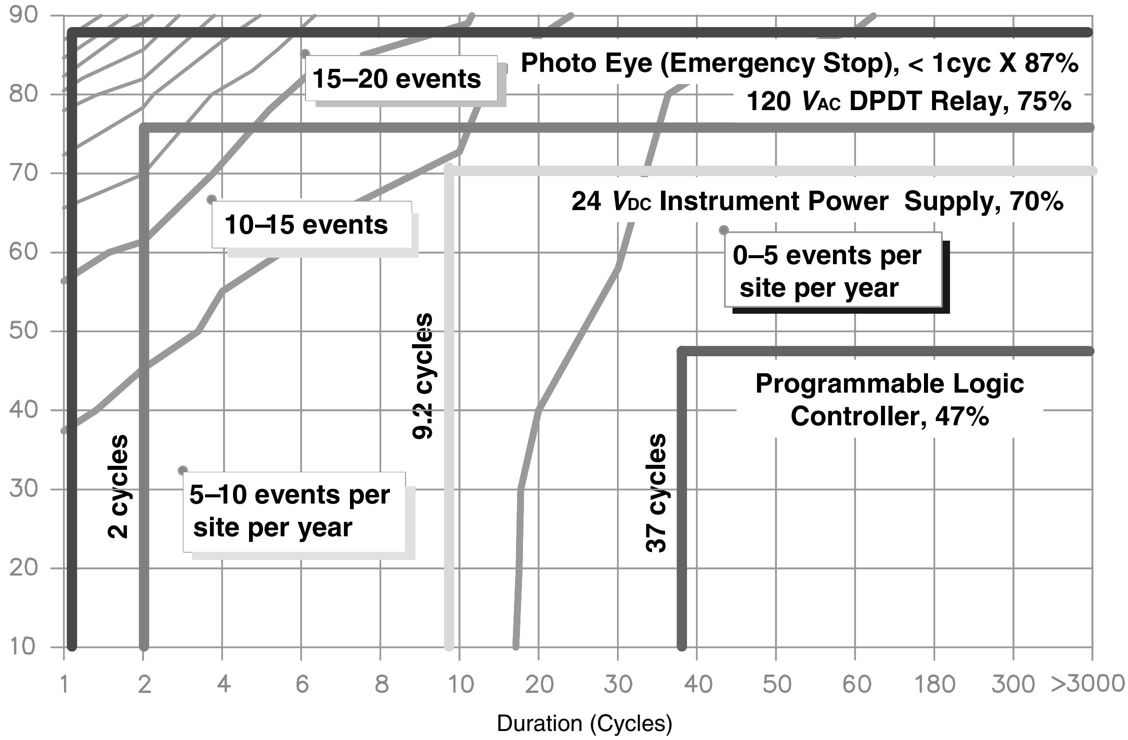


Fig. 5 Compatibility.

will result in process shutdowns each year. If the power supply were the weakest link in the process, a distribution served customer could expect between five and ten process interruptions a year. Unfortunately, the DPDT relay and the photo eye would cause far more disruptions.

Power Quality solutions can be applied at the facility level to protect the entire facility, at the process level to protect a process, at the equipment to protect the piece of equipment, or at the component level for example to hold in a relay or contactor. In general, power quality solutions applied at the facility level are the most expensive while solutions applied at the component level are the least expensive. There are situations however when each is appropriate. For example, if the cost of a process interruption is minimal, then perhaps improving the power supplies would be all that could be economically justified. However, if the cost of a process interruption is substantial, then a process DVR, dynamic voltage restorer, or perhaps even a flywheel would be the right business decision. Armed with this information and the cost of each solution, the customer can select and justify the appropriate level of voltage sag ride-through (Fig. 6).

TRACKING CUSTOMER REQUESTS

Each year customers contact their utility for help, resolving reliability or power quality problems. If the customer and the utility already have a good working relationship, the utilities are usually the first contacted. Sometimes the customer first contacts the equipment manufacturer and then at the recommendation of the manufacturer, contacts the utility. Unless there is an obvious problem with the equipment, the utility usually receives a call. This gives the utility the unique advantage of knowing what problems a particular business or industry is consistently experiencing. With this information, the utility can track and trend the data to determine what problems should be studied to find effective and economical solutions.

One tool used by utilities is the PQ Database. This database was developed by EPRI for utilities to use to store and sort power quality inquiry data. When a customer contacts a utility to request assistance with a power related problem, the description of the problem is entered into the database. As the information about the problem is gathered, it too goes into the database. Monitoring data is often collected and saved, and the actual resolution to the inquiry is also stored. This data can then be sorted in a variety of ways. Samples from a few such sorts are shown. Tables 1–3 illustrate such data collected by Duke Power's PQ engineers from 1997 to 2000.

Table 1 shows equipment affected by power quality events. Table 2 shows the responses of the equipment to the power quality events, and Table 3 shows the causes of the power quality events.

Once the data is collected and sorted, it quickly becomes obvious which equipment is experiencing the

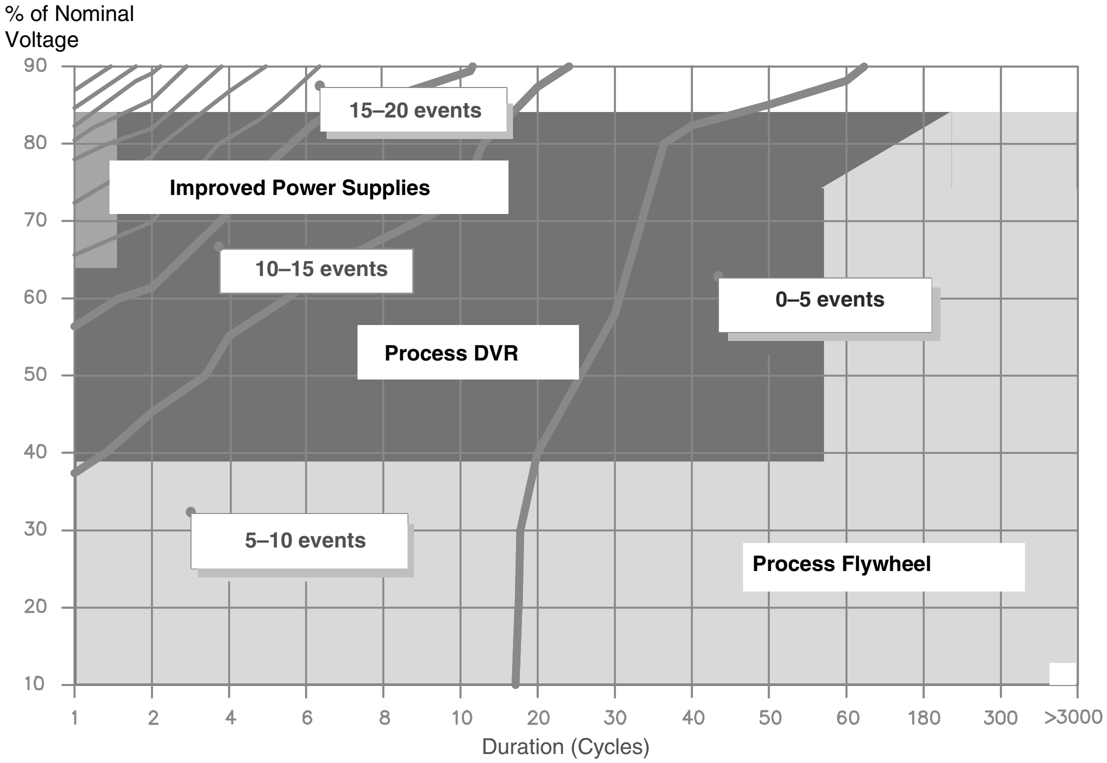


Fig. 6 Selecting a solution.

most problems and what problems are most often encountered. Many utilities contribute to a collaborative effort, to research ways to better solve common power quality problems. Often solutions exist, and it is just a matter of identifying the most appropriate solution and where the solution would best be applied. Sometimes existing solutions are not appropriate and a new solution is developed, applied, and tested. This effort requires collaboration between utilities, a research and development team, customers, and equipment manufacturers.

DEMONSTRATION PROJECTS

When ideal solutions to existing power quality problems do not exist, new solutions that solve the identified problems can be designed, developed, and tested. One example is the low speed flywheel. This flywheel was specified by EPRI, developed by Active Power, and installed and tested at Shaw Industries, a customer served by Duke Power. This demonstration project required collaboration and cooperation of EPRI, Active Power,

Table 1 Affected equipment from duke power PQ database

Affected equipment	Count	Percent
Plant equipment—in general	317	14.2
AC drive	204	9.2
Process controls	152	6.8
Motor	140	6.3
Other-affected equipment	97	4.4
PC	84	3.8
Breaker	73	3.3
LANS–Networks, Mainframes	62	2.8
UPS	61	2.7
Chiller, HVAC	61	2.7

Table 2 Equipment response

Affected equipment	Count	Percent
Equipment/component failure	501	27.5
Drop-out	334	18.3
Erratic operation	232	12.7
General-customer request	207	11.3
Nuisance tripping	154	8.4
Energy usage monitoring	65	3.6
Cycling—on and off	52	2.9
Lights dimming/flickering	42	2.3
PQ characterization	31	1.7
Fuse expulsion	29	1.6

Table 3 Problem causes

Affected equipment	Count	Percent
Duke system; fault, operation	209	13.4
Customer equipment degradation	155	9.9
No PQ problem	153	9.8
Other—customer system problems	148	9.5
Lightning strike	115	7.3
Grounding/bonding	94	6.0
Capacitor switching	91	5.8
Equipment specification	71	4.5
Problem cause no found	68	4.3
No electrical problem	61	3.9

Shaw Industries, and Duke Power. The flywheel was applied to a process line that extrudes filaments to produce carpet backing. This facility was served by a distribution circuit and the process shut down frequently due to voltage sags and momentary interruptions.

The key to successful implementation of the flywheel as a power quality solution included:

- Characterizing the electrical system,
- Determining the sensitivity of the process components,
- Matching the solution to the process line power requirements as well as to the utility distribution system relaying schemes,
- Including specific and accurate language in specifying the solution,
- Meticulously diagramming the process,
- Properly installing the solution, and
- Monitoring the results.

Shaw Industries realized both productivity and efficiency gains from the operation of the flywheel.

UTILITY/CUSTOMER FORUMS

Some utilities have customer groups that meet regularly to discuss power quality and reliability issues, as well as utility standards and changes, and customer or industry challenges. This forum offers a unique opportunity for the utility and the customer to find ways to help each other. Many power quality opportunities are discovered this way. Without this interaction, the utility could only guess what their customers needed and customers could only guess how and why utilities implemented policies and changes. One such group, the Power Quality Issues Forum was established at Duke Power in 1992. This group meets quarterly and participates in activities that include: visiting the Clemson University Electrical Engineering Laboratory to see utility funded power quality research and development efforts; attending PQA, a yearly power quality conference hosted by EPRI; traveling to EPRI-PEAC to see equipment and product testing and research; sponsoring power quality training to stay current on industry initiatives; and visiting demonstration sites such as the installation of the flywheel at Shaw Industries.

WORKSHOPS

As information is gained from demonstration projects, IDG work, SagGen testing, equipment testing done by EPRI-PEAC, and research and development conducted by universities and funded by utilities, this information must be shared with utility customers. One effective way of doing so is by holding workshops, conferences, and seminars. Some specific examples include the Industrial Technology Information Exchange conference sponsored by Duke Power yearly, for the past two years. This conference highlights new electric technology applications, power quality and reliability efforts, and innovative ways to reduce energy costs. Based on customer feedback from the first year, the conference was held twice in 2001 in both North and South Carolina and participation increased dramatically. Another successful way to share relevant power quality information is conducting workshops and seminars. Such workshops offered by utilities focus on wiring and grounding for enhanced equipment performance, effectively applying surge suppression, improving process ride-through, selecting effective power quality solutions, improving drive performance, selecting appropriate monitoring equipment for the application, designing effective lightning protection systems, and solving problems with ungrounded electrical systems.

TRACKING COSTS OF PQ PROBLEMS

In order for a customer to make an informed business decision about power quality solutions, it is imperative to know the cost to the business when a power quality problem occurs. This cost should include scrap, unproductive labor, lost productivity, missed orders, risk, and recovery. A process for tracking these costs should be implemented if not already in place.

SPECIFYING NEW EQUIPMENT FOR RIDE-THROUGH

In order to minimize costly power quality solutions, ride-through specifications can be included when purchasing new equipment. Ideally, one would characterize the

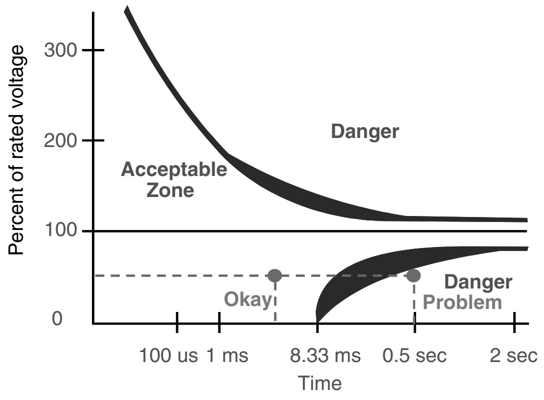


Fig. 7 The Information Technology Industry Council (ITIC) curve.

electrical environment to know exactly what specifications to include, but this is not always feasible. A generic ride-through curve can be included which would allow a process to ride-through most voltage sags. Fig. 7 illustrates the Information Technology Industry Council (ITIC) Curve. This curve was developed to represent the sensitivity of most computer equipment. The region within the two curves represents conditions where computer equipment should operate without disruption or damage. A similar curve has also been developed for the semiconductor industry.

KEEPING ABREAST OF NEW PQ SOLUTIONS

Utility customers can keep abreast of new power quality solutions in a variety of ways. Power quality magazines have articles describing new solutions. Utilities offer seminars and conferences outlining the features of power quality solutions. Literature from equipment testing is available through EPRI utility members. With the tendency for business and industry to reduce personnel, it is not always practical to stay informed about power quality issues. In which case, one can contact utility power quality experts when power quality problems occur or when specifying new equipment.

EDUCATING UTILITY ON PROCESSES

In order for utilities to help when power quality problems do arise, it is imperative that the utility power quality engineers be familiar with the processes in question. This can only be accomplished if the utility customer experiencing the problem feels comfortable sharing process information with the utility. Much information regarding processes is confidential and proprietary, and the utility must respect that confidentiality. It is much easier to expeditiously solve a power quality concern if the utility power quality engineer already has a good understanding of the process equipment. This requires resources and commitment from the utility management. Power quality expertise takes years to acquire technology, equipment, and power quality solutions are constantly changing, requiring time and effort to remain informed.

CONCLUSION

By actively participating in a partnership, both the electric utility and utility customers can significantly improve their business results. This requires collecting and sharing information regarding utility system operation, studying and sharing process requirements—including sensitivity to electrical disturbances, and evaluating solutions to power quality problems.

Information can be effectively shared through customer forums, conferences, and workshops. A utility, customer partnership requires commitment from all involved, but the benefits are well worth the effort.

BIBLIOGRAPHY

1. *Industrial Design Guide*: Version 3.0, EPRI: Palo Alto, CA, 2001; 1005912.
2. *Power Quality Applications Guide for Architects and Engineers*, EPRICSG: Palo Alto, CA, 1999; TR-113874.
3. *Duke Power Energy Solutions for the Information Age Workshop*, Charlotte: NC, May 23, 2002.

Util–Wall

Utilities and Energy Suppliers: Planning and Portfolio Management

Devra Bachrach Wang
Natural Resources Defense Council, San Francisco, California, U.S.A.

Abstract

Throughout the United States, most customers continue to receive service from their hometown utility companies, regardless of the status of retail competition in their state's electric industry. Recent turmoil within the industry has focused attention once again on a crucial responsibility of those utilities: electric-resource portfolio management. Effective portfolio management requires a fully integrated approach to identify customers' electric service needs and to select demand- and supply-side alternatives to meet those needs through a portfolio that minimizes the total cost and environmental impacts and has an acceptable level of risk. To enable effective portfolio management, regulators must align utilities' financial incentives with customers' interests. Traditional regulation creates a substantial financial disincentive for utilities to meet customers' energy service needs through cost-effective energy efficiency or other demand-side resources, but regulators and governing boards can and should eliminate this disincentive.

Every utility's current resource mix and its cost of providing energy is largely inherited from past portfolio managers' investment decisions. Long-term, integrated resource planning provides an opportunity for portfolio managers to work with their regulators or governing boards, customers, and other stakeholders in an inclusive and transparent manner to create a plan to meet future energy service needs. The resulting long-term plan should enable stakeholders to understand what the resource mix will be in ten or twenty years, the expected costs of the resource plan, and its environmental impact. Long-term plans provide a common basis for both the utility and the regulator upon which to found subsequent resource investment decisions.

The long-term planning process begins with a forecast of customers' demand and an accounting of the resources that are available to meet that demand. The portfolio manager should conduct a comprehensive analysis of the costs, risks, and environmental impacts associated with all resource options, including both demand-side and supply-side resources, that could be included in the portfolio to meet customers' needs. A number of potential resource portfolios should be tested against the portfolio manager's primary objectives. The final long-term plan should document this analysis and describe the elements of the preferred resource portfolio.

PROVIDING ENERGY SERVICES: THE ROLE OF UTILITIES IN SOCIETY

The overriding purpose of any utility or energy service provider is to meet its customers' energy service needs in an affordable, reliable, and environmentally sensitive manner (Throughout this entry, I often refer to the portfolio manager as a utility, since utilities are the most common entity to perform the portfolio management function; however, as I discuss in a later section, other entities also perform this function). While this may sound simple in concept, achieving these goals in practice is difficult and requires expertise in managing a diverse portfolio of resources to achieve sometimes conflicting objectives.

Managing a portfolio of demand- and supply-side resources is complex in part because the services that energy utilities provide customers vary widely. Utility customers do not purchase electricity from a utility for the sake of the electricity itself. Instead, customers pay the utility in order to receive the services that the electricity can provide. Customers want light to read by, cool and comfortable homes in the summer, and clean clothing, all for a reasonable price. As such, customers generally care more about the total bill they will pay to receive those services than about the rate for each kilowatt-hour of electricity consumed. In order to best meet customers' needs, utilities must focus on the broader scope of providing the energy services that customers want, rather than just the energy itself as a commodity.

Moreover, many of the services that utility customers demand, along with access to shelter, food, and water, are considered essential to achieve even a basic quality of life in modern times. As the single largest source of air pollution in the United States, the electricity industry must also be a key contributor in the effort to provide all people with another basic human right: access to clean and healthy air to breathe. Thus, how utilities provide their

Keywords: Portfolio management; Integrated resource planning; Risk management; Long-term plan; Resource procurement.

customers with energy services is fundamentally intertwined with a broad range of social and political issues.

THE BENEFITS OF PLANNING AND PORTFOLIO MANAGEMENT

Perhaps nothing in recent history demonstrated the need for long-term planning and portfolio management in the electric industry as clearly as the California energy crisis of 2001. (In 2002, the California Legislature enacted Assembly Bill 57, returning the utilities to the role of portfolio managers.[1] The California Public Utilities Commission has adopted several subsequent decisions providing guidelines for the utilities' portfolio management activities.[2,3]) That experience showed forcefully that simply relying on a spot market to meet customers' needs will not achieve the industry's overriding goals of providing customers with energy services in an affordable, reliable, and environmentally sensitive manner. To achieve these goals, utilities and other service providers must assemble a diverse portfolio of resources that will be robust in the face of the many uncertainties prevalent in the electric industry. Such a portfolio is likely to include demand- and supply-side resources with a mix of long-, medium-, and short-term commitments, owned and contracted resources, and supply-side resources with a variety of fuel sources, along with financial instruments to help manage any remaining risks.[4] Careful diversification along a number of different dimensions will help to avoid the classic pitfall of putting all the portfolio manager's "eggs in one basket" and can provide protection against risks, including those related to fuel prices, future loads, fuel supply availability, and future environmental regulations.

Who is Responsible for Portfolio Management?

Throughout the United States, the most common electric-resource portfolio manager is the hometown utility. This includes publicly owned utilities that are governed by local boards, as well as investor-owned utilities that are regulated by state Public Utilities Commissions. These boards and regulators are responsible for guiding the utilities' portfolio management and long-term investment activities.

In states that have restructured their electricity industry and established a competitive retail market, energy service providers (ESPs) also perform the portfolio management function on behalf of their customers, while the management of the distribution system remains with the utility. Most ESPs are required to sign up each individual customer that they serve; however, a few states, including California, Massachusetts, Ohio, and Rhode Island, have enacted legislation that enables local governments to become portfolio managers, commonly known as community choice aggregators (CCAs), by providing energy services to utility customers in their jurisdictions on an "opt-out" and aggregate basis.

The Long-Term Planning Process

Long-term planning is an essential responsibility of portfolio managers. A long-term view is necessary because the electricity industry is characterized by the need for capital-intensive investments with sometimes long lead times. These investments are often lumpy in nature, and portfolio managers' investment decisions must be made based on uncertain information. In addition, many new resources will continue generating or saving electricity for thirty to forty years or more, so the costs and benefits of investing in a particular resource must be analyzed over an extended time period. Moreover, the various resources from which a portfolio manager can choose to meet its customers' needs expose it and its customers to various types and levels of risks, which must be analyzed in the context of the overall portfolio of resources.

Portfolio managers should conduct integrated analyses of various resource portfolios and investment options on a regular basis, for example, every two to three years. These long-term plans can build off of the experience utilities and regulators gained with Integrated Resource Planning over the past quarter-century. Regular long-term planning processes enable a portfolio manager to compare resource alternatives in a manner that captures interactive portfolio effects. Without long-term, integrated planning, a utility that analyzes procurement options one by one is likely to "miss the forest for the trees." Each individual investment decision may seem like the best decision, but the *additive* effect of the decisions and the impact on the overall portfolio would not be considered without true long-term plans. The preferred resource plan generally has, among other factors, the lowest lifecycle cost (i.e., the lowest anticipated long-term revenue requirement) and is most robust in the face of various risks.

Long-term plans provide a portfolio manager and its regulators or governing board, customers, and other stakeholders with a common roadmap toward the future. Importantly, the process of developing a long-term plan provides a crucial opportunity for portfolio managers to engage all of these stakeholders in a collaborative, public process to understand the risks and benefits of alternative portfolios and to seek a general consensus on the best path forward. (Montana's guidelines for electricity procurement provide a good example of the policies governing, and the processes used for, portfolio management.[5]) Such a public process not only benefits the various stakeholders who have an opportunity to help determine the future of the portfolio, it may also help the portfolio manager to avoid potentially lengthy and adversarial regulatory hearings.

While long-term plans must delve into the technical details of load forecasts and the characteristics of resource alternatives discussed further below, a plan should also be able to answer key policy-level questions. For example, a long-term plan should enable policymakers and the public to understand what the portfolio manager's resource mix will be in ten or twenty years under its preferred plan and how it will differ from the present resource mix. The plan should also clearly present the expected costs of the resource plan (and the impact on average customer bills over time), the accompanying risks, and its impact on pollution emissions.

Long-term plans (or Integrated Resource Plans) serve as common guidebooks for both the utility and the regulator, so that subsequent resource decisions are founded upon common understandings and assumptions. These plans may assist utilities in making a strong case for cost recovery, and they help regulators ensure that the energy system is on an appropriate path toward the future.

UTILITIES' FINANCIAL INCENTIVES MUST BE ALIGNED WITH CUSTOMER INTERESTS TO ENABLE EFFECTIVE PORTFOLIO MANAGEMENT

Effective portfolio managers should analyze both demand- and supply-side resources to meet their customers' needs. In many instances, it is more cost-effective for a utility to help customers use energy more efficiently than to buy power (or build a power plant) to serve inefficient uses of energy. However, traditional regulation creates a substantial financial disincentive for utilities to meet customers' energy service needs through cost-effective energy efficiency or other demand-side resources.

This disincentive arises because once a utility's rates are set, its revenues are tied to the volume of electricity that is sold. If actual annual electricity sales diverge from the forecast that is used to set the authorized revenue requirement, the utility will either under- or over-recover the fixed-cost element of its revenue requirement. While this problem is most often discussed in the context of investor-owned utilities, publicly-owned portfolio managers, including municipal utilities and CCAs, can face similar "conflicts-of-interest" with demand-side resources if their governing boards use a similar process to set rates. Fortunately, this disincentive can be eliminated through the use of modest, regular "true-ups" in rates to ensure that any fixed costs recovered in kilowatt-hour charges are not dependent on sales volumes. This solution has been implemented successfully by a number of regulators to remove utilities' disincentives for investments in energy efficiency and other demand-side resources. (For a complete discussion of the conflicts of interest that portfolio managers can face relative to demand-side resources, and the solutions to it, see the reference section.[6])

In addition to eliminating these disincentives, development of the most reliable, affordable, environmentally responsible energy service portfolio requires, among other things, a balanced, performance-based incentive system that provides the portfolio manager with risks and rewards to the extent it achieves (or does not achieve) these objectives. More than a decade ago, the national association of regulatory utility commissioners (NARUC) urged its members to "ensure that the successful implementation of a utility's least-cost (investment and procurement) plan is its most profitable course of action."[7] The resolution framed the term "least-cost" over an extended time period. Congress endorsed NARUC's objective in the National Energy Policy Act of 1992, for both electric and gas utilities, although the final decision remains with state regulators.[8] NARUC's stated objective remains an important prerequisite to enable effective electric-resource portfolio management, although in most states, current regulatory incentives do not achieve this objective.

KEY ELEMENTS OF A LONG-TERM PORTFOLIO MANAGEMENT PLAN

The process of developing a long-term plan begins with a forecast of customers' demand and an inventory of the resources already at the portfolio manager's disposal to meet that demand. Next, the portfolio manager should conduct a comprehensive analysis of the costs, risks, and environmental impacts associated with all available resource options, including both demand-side and supply-side resources, that could be added to the portfolio to meet customers' needs. Finally, the portfolio manager should test a number of potential resource portfolios against its primary objectives, conduct a risk analysis under various scenarios, and assemble a plan for an optimal portfolio.[9]

Forecasting Demand and Existing Resources

Forecasting customers' energy service needs (both energy and peak-load demand) is the first step in long-term planning. The changes in demand over time are a function of many factors, including population, economic conditions, weather, and how energy is used by customers (known as "end-uses"). Portfolio managers can use both econometric and end-use forecasting techniques to project future customer demand.[10] Econometric models use a "top-down" approach to project future demand by extrapolating trends based on past relationships between demand and key metrics such as population, weather, and economic growth. End-use forecasting models produce "bottom-up" forecasts based on the penetration and energy consumption of various identifiable end-uses such as refrigeration, lighting, air conditioning, and other electrical equipment.

Together, these types of models can capture the key variables affecting changes in customer demand over time. However, since forecasts are inherently uncertain, portfolio managers should also conduct sensitivity analyses to understand the full range of possible future demand scenarios. These sensitivity analyses might, for example, analyze a low load-growth scenario under which population and economic growth estimates are at the low end of the range, and conversely, a high load-growth scenario with higher possible population and economic growth forecasts. In states with retail competition, these scenarios might also include forecasts of various levels of customers switching providers. This forecasting and scenario analysis process often results in a "jaws" forecast of future demand, defining both the best estimate of future demand as well as the potential range of future demand.

Once a portfolio manager has forecast the range of possible demand, the next step is to assess the existing resources that it either owns or has under contract, and to compare these available resources to the demand forecast. The difference between the two will determine the portfolio manager's remaining need for additional demand- or supply-side resources. During the planning process, the portfolio manager should also assess whether the portfolio could be improved by replacing or re-powering any of the existing resources.

Analysis of Resource Alternatives

To fill that remaining need, the portfolio manager should analyze the costs, risks, and environmental impact associated with the full range of potential resource options. These options include energy efficiency, distributed generation, renewable resources, thermal resources (such as natural gas-fired plants and integrated gasification combined-cycle coal plants), transmission, and more.

Energy efficiency is the most cost-effective, reliable, and environmentally friendly resource available to portfolio managers. Assessing the potential for energy efficiency resources requires an analysis of the various end-uses (i.e., how customers use energy), how much more efficient those end-uses could be, and what level of efficiency is achievable through voluntary programs that provide incentives and information to customers to improve their efficiency or through mandatory standards that set the minimum level of required efficiency. (California's recent analysis of the potential for cost-effective energy efficiency provides a good example of this type of study in potential.[11])

Determining what portion of that energy efficiency potential is cost-effective requires an analysis of the total cost to society of procuring the energy savings; the total resource cost (TRC) test is the cost-benefit test most commonly used by regulators and portfolio managers to determine what level of energy efficiency is cost-effective.[12] The TRC test accounts for the total cost to society of acquiring efficiency resources, including the incremental cost to customers (if any) of implementing more efficient technologies or practices and the cost to the portfolio manager of offering the programs. Since energy efficiency resources are comprised of many smaller resources, portfolio managers often shorten the analysis of determining what portfolio of programs is cost-effective by comparing the costs of the efficiency resources to a static estimate of avoided costs (including avoided costs of generation, transmission, distribution, and environmental pollution), rather than putting the energy efficiency programs into the iterative portfolio analysis alongside the supply-side resources.

Assessing supply-side options requires an analysis of the costs, attributes, and risks associated with each resource. Every resource's fixed and variable costs should be assessed either over the lifetime of the resource or over some fixed period, often thirty years. In order to allow all resources to compete on a level playing field in this assessment, portfolio managers must use accurate operating and cost assumptions for each resource. For fossil-fueled resources, forecasting fuel prices (with a sensitivity analysis) is a critical element of this cost assessment. In analyzing renewable resources, assumptions regarding capacity factors and integration costs must be realistic in order to accurately reflect the cost of these resources.

Different resources have different operating characteristics, which may make them more or less valuable to a portfolio manager. For example, some resources are designed to operate at a relatively constant output to serve baseload demand, whereas other resources are dispatchable and can ramp up and down quickly to follow load or to meet peak demand. The portfolio manager must consider these attributes in order to design a portfolio that enables it to maintain reliable service.

Each resource exposes the portfolio manager to certain risks while mitigating other risks.[13] For example, natural gas-fired resources expose the portfolio manager to fuel price risks, whereas renewable resources should help mitigate exposure to this risk in a portfolio. Conversely, some types of renewable resources are intermittent and expose the portfolio manager to reliability risks, whereas natural gas-fired resources are more dispatchable and can help mitigate this risk. Conventional coal-fired generation is particularly vulnerable to financial risks associated with the potential regulation of greenhouse gases. These are just a few examples; all of the significant risk attributes of each resource must be well understood so that the various portfolios of resources can be tested against the variety of risks and designed to minimize both risks and costs.

Finally, resources have widely varying environmental impacts. By analyzing the environmental profile of each type of resource, the portfolio manager can assess the projected environmental impact of various portfolio options to help select a portfolio that meets the objective of providing energy services in an environmentally responsible manner. This information is also necessary to assess the financial risk exposure due to pollution emissions, as we discuss further below.

Determining the Optimal Portfolio of Resources

The final steps in assembling a long-term plan are to test a number of potential resource portfolios to determine their total long-term costs, conduct a risk analysis of those portfolios under various scenarios, and select an optimal portfolio that best meets the portfolio manager's objectives. While optimization models have been used to help determine the preferred plan, most utilities construct dozens of potential portfolios "by hand." These portfolios should span a wide spectrum of possible options. The operation of these portfolios is then modeled over the timeframe addressed by the plan (often 30 years) and the portfolios' total long-term revenue requirements, environmental impacts, and other metrics are compared. As the analysis illuminates the components of the portfolios that are more or less desirable, new portfolios may also be constructed and tested.

This application of "base case" assumptions in the initial analysis should produce useful information, but given the numerous risks in the electric industry, it is essential to conduct a risk analysis to test how robust each portfolio is in the face of various uncertainties. There are generally at least three different types of risks: (1) risks that can be quantified and for which historical experience can inform assessments of the future risk (e.g., load forecasts, natural gas price risk, etc.); (2) risks that can be quantified but for which no historical experience can inform the assessment (e.g., future regulation of carbon dioxide emissions); and (3) risks that cannot be easily quantified, but can be assessed qualitatively (e.g., a change in FERC's market design, public acceptance of new resource siting, etc.) PacifiCorp's 2003 Integrated Resource Plan provides a good discussion of these different types of risks and develops a framework for analyzing the risks.[14]

Utilities have traditionally emphasized the first type of risk listed above in their analyses. However, the other two types of risks are no less significant or real; even if they can't be quantified based solely on historical experience, they can often be quantified and incorporated into the integrated resource analysis. The financial risk associated with future regulation of carbon dioxide emissions is a prime example of the type of risk listed in the second category above that utilities have historically failed to assess or mitigate.[15] As the electric industry becomes more sophisticated with risk management, a number of utilities are beginning to conduct comprehensive risk assessments. Leading examples include PacifiCorp, Idaho Power, Puget Sound Energy, and Pacific Gas & Electric.[16–19] The Northwest Power and Conservation Council is a leading public sector practitioner.[20]

Based on these cost, risk, and environmental performance analyses, the portfolio manager should select a preferred portfolio that best meets its objectives. The final long-term plan should outline the elements of that preferred portfolio, in addition to documenting the full analysis conducted to assemble the preferred portfolio.

TURNING A LONG-TERM PLAN INTO ACTION

The principal value of a long-term plan is determined by its translation into the actual procurement of resources by the portfolio manager. Long-term plans lay out the elements of a preferred portfolio over ten to thirty years or longer, but circumstances within the industry may change over time and affect the underlying assumptions that led to the selection of the preferred plan. As such, long-term plans should be living documents that are updated regularly, as needed. In the intervening years, the procurement of resources should be consistent with the long-term plan. However, if conditions have changed enough to warrant deviations from the plan, the long-term plan should provide a common basis for stakeholders and regulators to understand what has changed and how those changes affect procurement.

Different portfolio managers have utilized their long-term plans in resource procurement in varying ways. Some include a near-term action plan within the long-term plan itself, laying out the steps to be taken in the near future to work toward the long-term vision described in the plan. In some cases, utilities will build or contract for each specific type of resource described in the plan. In other cases, utilities will solicit competitive bids and determine whether to procure more or less of a certain type of resource based upon the offers that are received. Either way, it is important for the portfolio manager to remain flexible to respond to actual circumstances while still utilizing the guidance that is provided from the full portfolio analysis produced in the long-term plan.

CONCLUSION

Portfolio management is essential to enable utilities and other energy service providers to meet customers' energy service needs in an affordable, reliable, and environmentally sensitive manner. Long-term integrated resource planning is a crucial tool for portfolio managers to balance costs, risks, and other objectives in an industry that is characterized by widely varying resource options and long-lived investments. Recent turmoil within the electric industry has put a spotlight on the benefits of portfolio management and, in particular, the tools it provides for risk management. Regardless of what type of entity is responsible for meeting customers' energy service needs, or a region's electric industry structure, portfolio management and long-term planning enables all stakeholders to work together to create a common roadmap toward a better energy future.

REFERENCES

1. California Assembly Bill 57 (Wright, Chapter 835, Statutes of 2002).
2. California Public Utilities Commission (CPUC) Decision 03-12-062, December 18, 2003.
3. CPUC Decision 04-12-048, December 16, 2004.
4. Harrington, C.; Moskovitz, D.; Shirley, W.; Weston, F.; Sedano, R.; Cowart, R. *Portfolio Management: Protecting Customers in an Electric Market That Isn't Working Very Well*; Regulatory Assistance Project for the Energy Foundation and the Hewlett Foundation: Montpelier, VT, July 2002.
5. Administrative Rules of Montana, Default Electric Supplier Procurement Guidelines, Title 38, Chapter 5, Sub-Chapter 82, section 38.5.8201 et seq., December 31, 2003, http://161.7.8.61/38/38-6151.htm.
6. Bachrach, D.; Carter, S.; Jaffe, S. Do portfolio managers have an *inherent* conflict of interest with energy efficiency?. The Electricity Journal **2004**, *17* (8), 52–62.
7. Moskovitz, D. *Profits and Progress Through Least-Cost Planning*; The National Association of Regulatory Utility Commissioners: Washington, DC, 1989.
8. 16 USC section 2621(d)(8).
9. Biewald, B.; Woolf, T.; Roschelle, A.; Steinhurst, W. *Portfolio Management: How to Procure Electricity Resources to Provide Reliable, Low-Cost, and Efficient Electricity Services to All Retail Customers*; Synapse Energy Economics prepared for the Regulatory Assistance Project and the Energy Foundation: Cambridge, MA, October 10, 2003, www.raponline.org/Pages/Feature.asp?select=15.
10. Tellus Institute and Institute of International Education. *Best Practices Guide: Integrated Resource Planning for Electricity*. Prepared for United States Agency for International Development: Washington, DC. Available at: http://telluspublications.stores.yahoo.net/bespracguidi.html.
11. Rufo, M.; Coito, F. *California's Secret Energy Surplus: The Potential for Energy Efficiency*. Xenergy Inc. for the Energy Foundation and the Hewlett Foundation, 2002, www.energyfoundation.org/energyseries.cfm.
12. California Public Utilities Commission. *California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects*. October 2001, www.cpuc.ca.gov/static/industry/electric/energy+efficiency/rulemaking/eeevaluation.htm.
13. Bachrach, D.; Wiser, R.; Bolinger, M.; Golove, W. *Comparing the Risk Profiles of Renewable and Natural Gas Electricity Contracts: A Summary of the California Department of Water Resources Contracts*, LBNL-50965; Lawrence Berkeley National Laboratory: Berkeley, CA, April 2003.
14. PacifiCorp. 2003 *Integrated Resource Plan*, pp. 4–7, www.pacificorp.com.
15. Bokenkamp, K.; LaFlash, H.; Singh, V.; Wang, D. Hedging carbon risk: protecting customers and shareholders from the financial risk associated with carbon dioxide emissions. The Electricity Journal, **2005**, *18* (6), 11–24.
16. Idaho Power Company. 2004 *Integrated Resource Plan*, July 2004, www.idahopower.com/energycenter/2004irp.htm.
17. PacifiCorp. 2004 *Integrated Resource Plan*, 2004, p.155, www.pacificorp.com/Navigation/Navigation23807.html.
18. Puget Sound Energy. *Least Cost Plan*, April 2005, www.pse.com/about/supply/resourceplanning.html.
19. Pacific Gas and Electric Company. *Prepared Testimony*, R.04-04-003, Order Instituting Rulemaking to Promote Policy and Program Coordination and Integration in Electric Utility Resource Planning, July 9, 2004.
20. Northwest Power and Conservation Council. *The Fifth Power Plan: A guide for the Northwest's Energy Future*, Prepublication Draft, Document 2004-18, January 2005, www.nwcouncil.org/energy/powerplan/draftplan/Default.htm.

Utilities and Energy Suppliers: Rate Structures☆

Neil Petchers
NORESCO Energy Consulting, Shelton, Connecticut, U.S.A.

Abstract
This article discusses how electric, gas, and other regulated utilities charge commercial, industrial, and institutional customers for their energy services. In today's utility industry, there are numerous factors which make up a consumer's monthly electric or gas bill. This article contains sections that will individually describe each important component of most common electric and gas rate structures. A full understanding of each component will then allow a customer to comprehensively read and understand a monthly utility bill. Types of loads, cost allocation methods, billing factors, rate design strategies, common electric rates, and common gas rates will all be discussed.

INTRODUCTION

Natural gas and electric utility rates are determined through a regulatory process under the jurisdiction of a public utility commission (PUC). The utility periodically applies for a rate case hearing to request rate adjustments in response to changes in economic factors, such as supply, demand, distribution, and market forces that affect the utility's cost to serve. Rates are designed so that the utility can recover sufficient funds to cover costs and, in the case of investor-owned utilities, generate a reasonable return on investment for its stockholders. The PUC first determines the level of allowable investment by the utility in its facilities (referred to as the rate base) and its level of operating expenses. It then determines the rate of return the utility may earn on its investment and adds to it the operating expenses. This sets the total revenue requirement, which is equal to the total cost to serve. Rates are then designed to allocate revenue requirements among the various customer classes and subcategories within the classes.

Typically, customer classes include residential, commercial, industrial, institutional, and municipal. They may be subdivided into many rate classes or lumped together into broader rate classes. Commercial and industrial (C&I) as well as institutional (CI&I) customers, for example, are commonly lumped together.

General rate classes include categories, such as residential, small CI&I, and large CI&I. Residential rate classes are typically limited to single-family dwellings and multifamily dwellings metered separately from one another. Master-metered multifamily dwellings can be treated either as a separate rate class or as part of a commercial rate class.

Two common distinctions between customer classes are size of load and usage profile, although distinctions are made between CI&I classes based on other criteria as well. Time-of-use (TOU) and load factor are, in many cases, even more significant factors than size of load. Additional rate classes sometimes exist for cogenerators, nonutility generators, and other end-use- or equipment-specific categories. Another major distinction is between firm and nonfirm service. Many utilities are also able to offer special contracts that are customer-specific.

In assigning rate classes, utilities attempt to determine commonalties among customers within a class. This includes the assignment of load behavior. Contribution to a probability of a peak, for example, is an extremely important factor. Common characteristics within the class contribute to when the load occurs.

TYPES OF GAS AND ELECTRIC UTILITY LOADS

Consumer electric and gas loads may be characterized as peak, off-peak, baseload, and seasonal. Each is rated differently by the utilities, produces different costs to the end-user, and has different implications for the system.

PEAK AND OFF-PEAK LOADS

The broadest and most critical distinction is between peak and off-peak loads. Peak loads are those that occur during the time periods in which the utility must supply the maximum amount of energy and, therefore, use the highest amount of system capacity. Increases in peak load are

☆ This entry originally appeared as "Utility Rate Structures" (Chapter 21) in *Combined Heating, Cooling & Power Handbook*, by Neil Petchers. Published by Marcel Dekker, Inc., New York, NY, 2003; reprinted with permission from AEE/Fairmont Press.

Keywords: Utility rate structures; Utility rates; Electric rates; Gas rates; Seasonal rates; Peak load rates; Customer charge; Energy charge; Demand charge; Power factor.

closely tied to the need for investment in facility expansion. For electric utilities, peak loads often necessitate the use of the less fuel-efficient peaking plants. During peak periods, the rate of transmission and distribution system line losses also increases, further adding to supply requirements. For these reasons, peak capacity is the most expensive to purchase and carries with it the burden of increased capital cost and decreased fuel and delivery system efficiency. For gas utilities, peak loads that must be served on a firm basis determine the capacity requirements of the local distribution system piping network, the amount of gas supply and transmission capacity that must be reserved on interstate pipelines, and/or the amount of gas storage capacity required.

Off-peak loads are those that occur during the time of day, week, month, or year when the utility uses a relatively small amount of its total system capacity. Consequently, off-peak loads do not contribute to the need for facility expansion. For electric utilities, off-peak loads are usually served by the most efficient electric generation plants and distributed with relatively low line losses. For gas utilities, off-peak loads are generally served by a low-cost source of gas supply. For all of these reasons, off-peak capacity is the least expensive to purchase and carries with it the benefits of operating cost-efficiency and minimized capital cost impact.

In addition, many utilities further segment loads into intermediate categories, sometimes referred to as shoulder- or intermediate-peak periods. These loads and their effects lie between the peak and off-peak loads.

BASELOADS

Baseloads, from an end-user perspective, are those that occur all of the time (i.e., a 100% load factor). The quantity that can be referred to as baseload is defined as the baseload rate times every hour in a given period (i.e., day, month, and year). Baseloads are served, in theory, by the portion of the utility's capacity that is used constantly. Baseloads are often considered optimal loads because the fixed costs, associated with capacity-related investments, are spread over the maximum potential units of sales, and, in the case of electric utilities, the life-cycle costs of highly efficient power plants are at the lowest possible rate.

In many rate structures, baseloads carry a higher weighted average cost than off-peak loads. This is because off-peak load is a component of baseload that is separated from the peak component. Capital investment-related costs are frequently stripped away and added to the peak constituent. Variable operating costs, which are typically lower than the average due to increased efficiency, are broken out and assigned incrementally to off-peak load. Baseload costs are thus a composite of a fixed load experienced continuously in both periods.

Different rate structures separate or stratify costs to varying degrees between peak and off-peak. Traditional rate structures tend to have a relatively low differentiation of costs in different rate periods. On the other hand, more progressive rate structures have a wider range of prices. They attempt to assign costs to a greater number of finite blocks of usage based on their varying impact on capital and operating costs, with the result that the cost of a kilowatt-hour (kWh) of electricity or thousand cubic feet (Mcf) of gas at any given time may be far different than the average cost.

SEASONAL LOADS

Utility systems are generally either summer or winter peakers, referring to the season during which their peak-capacity requirements are highest. Most of the nation's electric utilities are summer peakers, because summer cooling loads outweigh the winter-heat-load component. In most regions of the country, heating loads are mostly served by direct on-site fuel use, such as gas or oil.

Even though fuel costs (driven by supply and demand) are even greater in the winter, the capacity-related capital cost component and the inefficiency of electric generation peaking plants produce a greater effect on seasonal electric costs than do their fuel cost component. Summer peakers, therefore, will often construct seasonally differentiated rates that are higher in the summer than in the winter.

However, not all electric utility systems are summer peakers. Some are winter peakers, some are balanced, and some (due to a high growth rate) set a new peak every season. Winter peakers are found in certain northern regions where cooling loads are modest and/or electric heat is predominant. Winter peakers or balanced systems may also be found in moderate climatic regions that are more conducive to the use of electric heat. Balanced systems are more likely for utilities with a heavy industrial base in which temperature-related end uses are minor, compared with the baseload process end uses.

Almost all gas utilities are winter peakers, due to the preponderance of space heating loads. Northern climate local distribution companies (LDCs) tend to have the most dramatic winter peaking load profiles. The LDCs that serve a significant industrial sector and/or provide large quantities of gas to electric generation plants tend to have more balanced loads, as do LDCs in warmer climates.

LOAD FACTOR AND USAGE PROFILE

The load factor is one of the most significant elements of a customer usage profile and utility's operating profile. For the utility system, it is the measurement of actual energy output over a period vs potential output over that period, based on the full capacity of the system. If, for example, an electric utility has 10,000 MW of total capacity and generates 120,000 MWh over a 24-h period, its load factor is 50% [120,000 MWh/(10,000 MW×24 h)]. If this utility had a 100% load factor, it could generate the same 120,000 MWh over 24 h with only 5000 MW of capacity. Hence, an incremental load with a 100% load factor will produce a significantly lower revenue recovery requirement than one with a 50% load factor, because the capital cost associated with the additional 5000 MW of capacity is eliminated.

If, for example, a gas LDC has a maximum daily distribution capacity of 100,000 Mcfd and a 100% annual load factor, it would distribute 36.5 million Mcf (100,000 Mcfd×365 days) per year. If, in actuality, it distributed only 14.6 million Mcf per year, the annual distribution system load factor would be 40% (14.6 million Mcf/36.5 million Mcf).

A load factor of 100% is the theoretical ideal operating state for utilities. In this ideal state, the utility's investment in capacity is spread over the maximum amount of potential output, and the revenue requirement per unit (kilowatt-hour or thousand cubic feet) sold is minimized. In reality, however, a 100% load factor is not an attainable goal. Some margin of excess supply, transmission, and distribution capacity is necessary for maintenance and emergency backup. It is also not possible to balance consumer loads perfectly.

There are, however, some conditions that produce utility load profiles approaching a 100% load factor. For example, electric utilities that operate as part of regional power pools and/or have sufficiently low operating costs to allow for export of all excess capacity can approach a 100% load factor. Utilities that are capacity constrained may also need to operate all of their plants at 100% load factor, and purchase power to meet the rest of their load requirements. Gas utilities that predominantly serve heavy industrial loads, for instance, may have a high load factor.

In order to improve system load factor, utilities often seek to market electricity or gas in low-load periods at a relatively low cost. They may also provide incentives, through various conservation and load management programs, to promote elimination or shifting of peak loads.

The utility considers load factor both on a discrete minute-by-minute basis and an hourly, daily, weekly, monthly, and yearly basis. Other common measurement periods are seasons, normal workdays, workweeks, and weekends. Sophisticated dispatch modeling is used to determine when to bring additional capacity (i.e., electric generation plants or gas storage) on- and off-line in response to load fluctuations.

Electric utility systems (or regional power pools) have a dispatch stack order. Generating stations are arranged in the order in which they will be brought on- and off-line in response to changing load requirements. Typically, plants in the stack are classified as baseload, intermediate, and peak-load plants. Baseload plants are generally the most cost efficient to operate, or have the lowest operating cost, including fuel costs per unit output. Intermediate (or swing-load) plants run much of the time, often under varying loads. Peak-load or peaking plants are typically the least costly plants to construct, but also the least cost-efficient to operate. The LDCs also have a type of dispatch order. Supply sources, including storage reserves and liquid natural gas facilities, are arranged in the order in which they will be used.

Load factor is one of the most important determinants of the cost to serve a given facility. Consider two electric utility customers that have the same monthly load of 72,000 kWh. One of the customers uses 100 kWh every hour of the month (720 h). Under a typical demand/commodity-type rate, this customer would have a peak demand of 100 kW and a monthly load factor of 100%. The other customer uses 75 kWh every hour of the month, except for 1 h every day, when a certain process requires 600 kWh. This customer would have a peak demand of 600 kW and a monthly load factor of 17% [72,000 kWh/(600 kW×720 h)]. While, in this example, the monthly usage is exactly the same for both customers, the cost to serve the customer with the 17% load factor would likely be several times that of the customer with the 100% load factor. To serve the customer with the 17% load factor, the utility would have to reserve 600 kW in generation, transmission, and distribution capacity as opposed to 100 kW for the other customer. If both customers paid the same amount for electricity on a per kilowatt-hour basis, much of the charges paid by the customer with the 100% load factor would be to support the investment required to serve the other customer.

Now consider a third customer that uses 72,000 kWh/month. This customer operates a night and weekend shift factory that uses 200 kWh every hour for half of the hours every month (360 h). Therefore, this customer has a maximum demand requirement of 200 kW and a monthly load factor of 50% [72,000 kWh/(200 kW×720 h)]. In this case, since the entire load occurs in the utility's off-peak period when load requirements are low, the utility does not require any additional capacity to serve this customer, except for the local wires and transformers connected directly to the facility. Additionally, this customer's loads can be served by the utility's most efficient generation plant. When the power is required by the customer (i.e., the specific usage profile), in this example, it is even less costly to serve the off-peak customer with the 50% load factor than it is to serve the customer with the 100% load factor.

A parallel example can be drawn with three gas utility customers that all consume the same amount of gas on an annual basis. The first customer has a 17% load factor, based on a winter heating load requirement. The second customer has a 100% load factor, based on a continuous industrial process requirement. The third customer has a 50% load factor, based on a continuous non-winter-month gas-fired cooling load requirement. Similar to the three electric utility customers, in this hypothetical example, the customer with the 17% load factor will likely be the most costly to serve and the customer with the 50% load factor will likely be the least costly to serve, with the 100% load factor customer falling somewhere in the middle.

The conclusion that can be drawn from these examples of customers with the same exact daily, monthly, or annual usage is that the load factor and, even more importantly, the specific load profile are key determinants of the actual cost of service. The main reason is the impact of the load profile on fixed cost requirements and how fixed costs are recovered with respect to overall usage requirements.

In addition to load profile, the magnitude of overall load is an important cost factor. There is an economy of scale in serving customers that requires large amounts of energy. Costs of running lines or pipes to a facility, installing service and meters, billing every month, and providing other goods and services are less when averaged over a large, rather than a small volume.

ALLOCATION OF COSTS

Utility allocation of fixed costs and the associated setting of rate levels is a process that is partly scientific and partly subjective. One major consideration in cost allocation and rate design is the desire for rate stability. All usage charges should contribute to fixed costs in some way, and demand charges should serve to mitigate against significant distortions in cost allocation resulting from varying load profiles.

If a utility placed all capital cost recovery in a single period of use, such as peak summer, the rate might be too unstable. The market might overreact with a rush toward solutions, such as peak shaving and, in short order, cost recovery would be insufficient. If all usage in that single usage period disappeared, the capital costs would not disappear. The utility system would still have to be financially supported. Of course, loads do not disappear all at once. Hence, as they change, the utility rate structures must change with them.

Fixed cost-related charges are neither arbitrary nor capricious, but they are also not a perfect scientific cost allocator. The underlying theory behind demand charges and other fixed cost rate components is that they better approximate real costs. If, for example, all customer electric bills under demand rates were broken down into incremental usage costs, the cost per kilowatt-hour might range from $0.02 to $0.90, or even higher, depending on the severity of the impact of demand charges. This wide range of price differentiation is said to allow rates to more closely represent the utility's actual cost to serve. As this process continues to evolve, along with advances in measurement and market communication technology, such determinations may be made dynamically, based on actual events, as opposed to predicted events.

Clearly, the movement in today's energy market is toward more precise cost allocation strategies. Factors that have contributed to this movement include the following:

- The need to protect utility revenues by having rates that are competitive with energy alternatives. More competitive, diversified rates have been constructed by electric and gas utilities to protect revenues and maximize sales during periods when electricity or gas is less costly.
- The desire of utilities to influence customer usage patterns in order to create higher system load factors. Price differentiation directs equipment investment choices and operating schedules toward customer load profiles that improve system load factor and lower the average cost to serve. If all similar energy units cost the same, load growth would tend to move away from a high load factor as the prevailing forces of weather and the normal work week directed loads to peak periods. Cost differentiation provides incentives to counter these forces through careful end-use planning.
- The directive from regulators to redesign cost allocation processes and avoid the need for additional capacity. Careful review of utility management decisions by regulators has directed utilities away from the business of building new plants much in advance of load growth.

In addition to least-cost planning, regulators have demanded more precise cost allocation in an effort to make rates more fair. Regulators have also provided utilities with incentives for activities other than construction. A preferred rate of return for investments in conservation and load growth reduction activity is an example. Many demand-side management programs involve shifting usage from the utility peak period to an off-peak period. In order to make such programs effective, utilities must have rate structures that charge according to when energy is used instead of just how much energy is used.

The imperative of the regulated least-cost planning activities has increasingly been replaced by competitive market-based imperatives. As a result, utility rate design is driven more and more by competition and less by static regulatory planning.

On the other end of the spectrum from a single fixed-price usage charge is a completely market-driven discrete usage charge that changes from hour to hour, or even minute to minute. In this scenario, gas and electricity usage charges continually vary with the price being established on almost an instantaneous basis. This price must be market competitive and reflect all embedded fixed and variable costs. This type of pricing would be applied to both the commodity side and the transmission and distribution sides, either separately or as a bundled price.

While the market is still not fully mature enough or unencumbered by imperfection to function with such pricing, the technology of metering and transmitting gas and electricity is approaching the level of sophistication necessary for such a market. Moreover, the forces of competition are moving the market in this direction.

However, today's market does not yet function in this manner. Instead, some utility rates are designed to approximate this type of instantaneous or real-time pricing (RTP). To do so, rate designs use a mix of various rate components. They can be classified into several general categories on the basis of energy-use characteristics. Each category may constitute one rate or several categories may be aggregated to determine a rate.

BILLING FACTORS

Utilities use a number of billing factors in their rate structures. Some of these factors are clearly stated in all utility bills, while some are considered in every transaction but not necessarily identified separately in all bills. Others are considered and applied, depending on the particular characteristics of the utility or the customer class under a given tariff. The most common billing factors are as follows:

- *Basic, or customer, service charge*. A fixed charge is assessed to each customer based on costs related to connection, metering, billing, service maintenance, etc. Typically, basic service charges are greater for rates designed to serve large users because of the greater cost associated with larger pipes, regulators, and metering equipment. Facilities with multiple services under the same rate may have summarized billing with only one basic service charge. Facilities with multiple services under different rates will often pay a basic service charge for service under each rate.
- *Minimum charge*. This is the lowest bill a customer is required to pay for service on a given rate schedule during each billing period, regardless of actual usage. In most cases, this is equal to the customer charge. However, for larger customers, it can include other charges, such as demand charges or charges established by individual contracts between the customer and the utility.
- *Commodity charge*. This is a charge based on energy usage or the number of energy units actually consumed by the facility during each billing period. Commodity charges generally include the utility's incremental operating costs, plus some contribution to fixed costs.
  - — Common natural gas billing units are: cubic feet (cf), hundred cubic feet (Ccf), thousand cubic feet (Mcf), and therm (100,000 Btu or 105,480 kJ). While a therm is a specific quantity of energy, a cubic foot of gas has varying British thermal unit or kilojoule levels. Typically, 1 cf of pipeline quality natural gas contains about 1000 Btu, but this can vary by several percentage, depending on the specific source of natural gas.
  - — Common electricity billing units are: kilowatt-hours and kilovolt-amperes.
- *Demand, or maximum level of service, charge*. This charge is based on a customer's peak rate of consumption and takes into account the utility's required investment needed to serve that load. Measurement of demand, or rate-of-use, is somewhat like a car speedometer. However, a demand meter does not measure rate-of-use instantaneously, but averages it over a discrete utility-selected interval. This interval, or "demand window," for electric utilities is usually 15 or 30 min. The interval for gas utilities is usually the highest daily consumption or, in some cases, it may be measured as the peak usage for 1 month. Demand can be measured in multiple-use periods, and charges can be differentiated for each use period.

Demand charges are assessed in several different ways. They may vary by season, TOU, and level of use:

- — Seasonal variation produces greater charges for peak demand during months in which the utility experiences its highest peak demand. With gas utilities, demand charges are often applied only in winter months.
- — The TOU variation can be applied in several different ways. In many cases, demand is only measured by electric utilities during a peak period. In other cases, demand is measured during several different periods, such as peak, shoulder, and off-peak. The charge per unit of demand varies with each period, peak charges being the highest.
- — Level-of-use variation produces different charges for increasing blocks, or steps, of demand. For example, the first 500 kW or 500 Ccf of demand will be billed at one rate, the next 1000 kW or 1000 Ccf of demand will be billed at another rate, and so on.
- — Minimum demand charges are sometimes set at a given level for each rate tariff. In some cases, they may be based on a percentage of the customer's connected load or main transformer or gas meter size.

- *Ratchet adjustments*. A demand ratchet adjustment sets minimum monthly billable demand at a certain percentage (usually 60%–100%) of the highest month's peak demand in a given preceding period (typically 11 or 12 months). When the utility has a severe summer or winter peaking system, the ratchet may be based on the highest month's peak demand only in the peaking season. The minimum monthly demand charge assures the utility of cost recovery for investment in peak capacity, even if a customer does not require that peak capacity in a given month. In some cases, this preceding period may be as long as several years or for the life of a contract. Ratchets serve as a mechanism to approximate the true cost of service and to distribute that cost impact on the customer over several months.

Consider the example of an electric utility with an 80% demand ratchet adjustment. If 1 month had a peak demand of 2000 kW and the next 11 months had a peak demand of 1000 kW, each of the proceeding 11 months' demand charges would be based on 80% of the 2000 kW figure or 1600 kW. The result would be an additional annual billable 600 kW/month, totaling 6600 kW additional billable kilowatt over the next 11 months. Fig. 1 provides an example of the impact of an 80% ratchet on an 11-month rolling basis.

The reasoning behind ratchet adjustments is that monthly demand charges alone do not sufficiently compensate the utility for the impact of a customer who sets a large peak once per year rather than every month. The utility must have sufficient capacity to meet the annual highest peak hour of demand at all times and must incur the cost of building or securing contracts to purchase the needed capacity. A customer does not need that peak demand every month must still pay for a large portion of the demand in order to reserve it for the peak month.

- *Energy adjustment charge (EAC).* This mechanism is designed to pass increasing or decreasing fuel costs per billable unit directly to the customer. Energy adjustment charges may vary on a monthly, quarterly, or annual basis. It is normally based on a 12-month rolling average of fuel costs and used to maintain rate stability between rate-case proceedings. Common terms used to express EAC are: fuel cost adjustment, fuel adjustment charge (FAC), purchased power adjustment, purchased gas adjustment (PGA), and levelized gas adjustment. Fig. 2 is an illustration of a FAC to monthly electricity charges.
- *Other adjustment charges.* A utility bill may also include additional adjustment charges. Many electric companies have a nuclear capacity adjustment or nuclear plant decommissioning charge. Some companies also have conservation adjustment mechanisms or sales adjustments. Gas utilities have shrinkage or retainage charges. These are applied as a percentage increase to transported gas volumes at delivery points for transportation customers. This charge reflects lost or unaccounted for gas volumes that arise from the transportation of gas over the LDCs pipeline network.
- *Taxes and fees.* These are added charges that the utility bears as operating expenses and passes on to individual customers. These may include gross receipt taxes or sales taxes. They may also include surcharges for special regulatory commission approved programs that are added to the bill. Such charges may be shown on the bill as itemized charges, or they may be embedded in the derivation of other utility charges and are thus not easily spotted on the bill.

RATE DESIGN STRATEGIES

The purpose of rate design is to recover revenue requirements associated with the costs incurred by utilities to provide services to the customer classes, while also recognizing the different energy use profiles of customers between and within the customer classes. Rate design also sends price signals to consumers and can influence energy-use decision makers to favor more cost-efficient energy use profiles. Price signals mean that a particular utility's rate design makes it clear when it is more or less costly to purchase a unit of energy.

If all customers had the same usage pattern, fixed and variable costs could be readily integrated into one simple cost recovery mechanism—a usage charge. To determine an average price per kilowatt-hour or thousand cubic feet, the total revenue requirement would be divided by all of

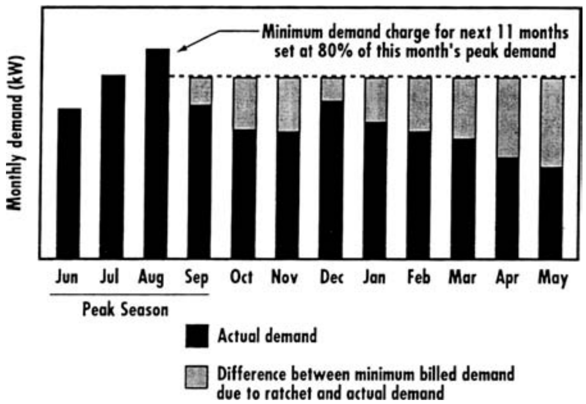


Fig. 1 Impact of 80% demand ratchet.

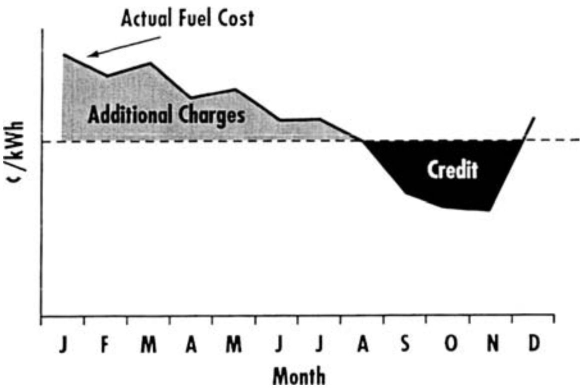


Fig. 2 Representative illustration of fuel adjustment charge to monthly electricity charges.

Util–Wall

the kilowatt-hour or thousand cubic feet to be sold. In reality, however, customers do not have the same usage patterns and, as demonstrated in the above load factor examples, distinctions must be made as to how varying customer load patterns affect system cost.

BLOCK RATES

Block rates are rates in which the charges for a unit of service vary with consumption. The billing period's consumption levels within a rate are often broken down into blocks, or steps, with different charges for each block. There are several different common types of block rates:

- *Declining block rates*. Historically, the most common type of rates, declining block rates have lower usage charges as levels of consumption increase. The reasoning behind these rates is that increasing customer usage reduces the cost to serve the customer on a per-unit basis. These rates have been phased out in many utilities because they are believed to encourage greater consumption and discourage conservation. Fig. 3 illustrates a natural gas declining block service rate.
- *Increasing, or inverted, block rates*. These rates charge more per unit as levels of consumption increase. They are relatively uncommon with natural gas service, but they do show up with electric services. They are used by utilities that are supply-constrained, or used for conservation purposes to discourage increased consumption. Fig. 4 illustrates a natural gas inverted block service rate.
- *Sliding block rates*. These rates use a peak demand value times a multiplier to determine the first-step size. They encourage greater load factors or, in other words, more constant levels of energy usage by the customer. This type of rate is common today in the electric industry for C&I customers. Fig. 5 illustrates a natural gas sliding block service rate.

In some cases, the utilities charge the same unit price regardless of the level of consumption, meaning there is only one block. This is somewhat common for residential customers.

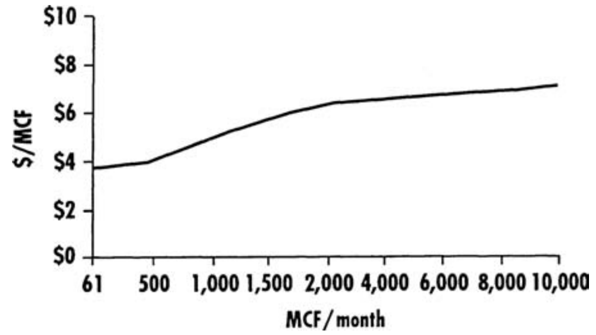


Fig. 4 Representative natural gas inverted block service rate.

Many utilities use a combination of block-pricing structures. They are often used to reflect a seasonal energy-cost differential. A utility may use an inverted block rate in the season of highest consumption and a flat or declining rate in the off-peak season.

DEMAND/COMMODITY RATES

These rates consist of two basic components: a demand charge and a commodity charge. The demand charge is typically based on the peak hourly or daily volume in the billing period. The commodity charge is based on the units of energy consumed. In many cases, a ratchet penalty is used. As previously discussed, the ratchet penalty is typically based on a certain percentage of the peak demand incurred within the previous year on an 11- or 12-month rolling basis. This type of rate somewhat more accurately reflects the cost to serve than flat commodity rates, because it takes into account peak-day requirements. Such rates encourage customers to maintain a high load factor.

In many cases, a sliding block rate structure is used to blend demand and usage charges into stepped rates based on utilization (or load) factors. Usage is billed at varying

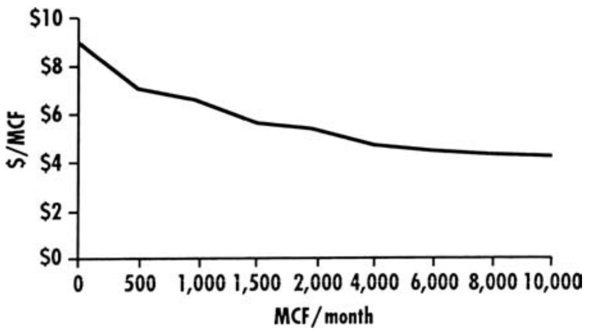


Fig. 3 Representative natural gas declining block service rate.

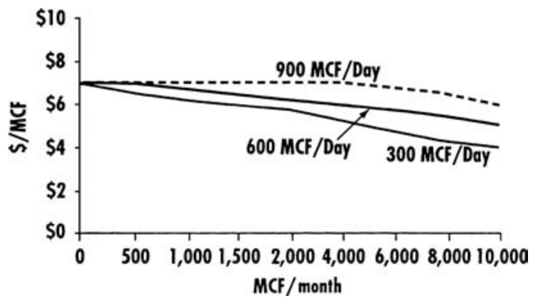


Fig. 5 Representative natural gas sliding block service rate.

Util–Wall

rates based on hours of use of demand. For example, a 30-day billing cycle has 720 h. If the peak demand on an electric utility bill is 1000 kW, a 100% load factor would represent 72,000 kWh, or 720 h of use of demand. A 25% load factor would represent 18,000 kWh, or 180 h of use of demand. This type of rate schedule breaks the hours of the month into different blocks. With a declining block rate structure, in each subsequent block or hours of use of demand, the rate is lower. For example, the first 180 h of use of equivalent full-load demand (18,000 kWh) would be billed at a certain rate per kilowatt-hour. The next 180 h of use of demand (from 18,000 to 36,000 kWh) would be billed at a lower rate per kilowatt-hour, and so on. This type of rate rewards high-load factor customers. If the customer's peak demand increases, more kilowatt-hour are billed at the higher rate and vice versa.

TIME-OF-USE RATES

The TOU rate structures are often designed to differentiate among months of the year (seasonal rates), days of the week, or hours of the day. These rates assign greater costs to peak usage periods, to discourage consumption, and lower costs to off-peak periods. Commodity cost differentiation may be augmented by peak demand charge differentiation. These TOU rates are more reflective of the cost to serve than standard block rates and provide price signals that direct consumers toward off-season or off-time usage. The level of differentiation will often be greater for utilities with poor annual load factors, such as those with loads that are predominantly heating or cooling. Typically, higher summer rates will be used by electric utilities and higher winter rates will be used by gas utilities.

- *Seasonally differentiated rates*. These are the type of TOU rate designed to assign greater cost to consumption, on a per-unit basis, for corresponding rate blocks in peak usage months. Commodity cost differentiation may be augmented by peak season demand charges and ratchets. These rates provide price signals that influence the consumers toward off-season usage.
- *Off-season rates*. These services are provided to customers in specific nonpeak months. They are designed for customers who do not require gas service in peak months, but do not have the alternative energy sources typically required for standard interruptible sales service. Similar to interruptible rates (IR), which are discussed later, these rates provide lower cost gas, because they do not contribute to the high cost of maintaining peak facilities. These rates may also be attractive to customers that have alternative energy sources, but cannot use them in certain months because they must meet seasonal emissions standards.

END-USE RATES

End-use rates are designed to provide incentives to customers to install and operate specific equipment that uses the utility's energy. This usually includes space heating, water heating, and cooling equipment. These rates generally have some basic equipment requirements and often need separate metering. Specialized heating rates are generally used by electric utilities. These are sometimes combined into an all-electric rate that is offered to customers using electric heating, water heating, air conditioning, cooking, etc. Specialized cooling equipment rates are more commonly offered by gas utilities.

End uses selected for specific rates can often be placed in given rate periods or seasons. In some cases, end-use rates are combined with TOU rates. The lowest rates typically are offered for end uses employed during the off-peak season or off-peak time of day or week. More moderate rates may be offered for baseload process end uses, such as water heating, cogeneration, or year-round industrial processes.

Many PUCs prefer that price signal tools be applied uniformly to usage characteristics rather than to specific end uses. However, some utilities feel that their typical rates already accomplish this and that further distinctions are needed to attract (or discourage) certain loads.

NONFIRM SERVICE RATES

Nonfirm rates include IR, standby rates, and load management rates. With multiple energy source options increasingly available, nonfirm service is becoming increasingly popular:

- *Interruptible rates*. These rates, which are commonly offered to CI&I customers, are designed for customers whose entire load, or large blocks of load, can be dropped at virtually any time. The primary benefit to utilities is that it reduces the need to guarantee service during peak demand periods. The utility can pass on savings, which come predominantly from reduced fixed-capacity costs, to customers. These rates are also used to attain a competitive advantage and/or attract loads based on lower costs. Interruptible service may involve commodity or distribution components.

  There are many conditions under which a facility can withstand such interruptions. Examples include:
  - Customers with dual-fuel burners capable of operation on natural gas and alternative fuels, such as propane or oil, or distribution systems supplied by propane–air mixtures can easily withstand interruption of natural gas service by switching over to their alternative burner fuel.
  - Customers with on-site electric power generation capacity can go off-line and generate their own

power to serve the entire facility or selected circuits within the facility.

— Customers with equipment that can operate on either electricity or fuel or steam, can simply use the nonelectrical equipment during periods of interruption. This would be the case with a dual-drive mechanical service device that had both an electric motor drive and a prime mover drive, or with mixed energy source (hybrid) multiple unit systems that feature both electrical and nonelectrical powered units.

Electric utility IR are often referred to as utility-controlled peak shaving. Customers are required to reduce their demand on the utility system completely, or to some predetermined level when necessary. Contracts may be designed with a set rate-break on demand or usage, a flat annual or monthly fee paid by the utility, or a specific payment rate for each period of interruption. Significant penalties for failure to interrupt are also common.

The LDC IR are typically based on the customer's ability to use an alternative fuel, such as oil or propane with the natural gas commodity charge set or negotiated based on the price of the alternative fuel. Natural gas distribution services (in cases where a customer purchases gas from a seller other than the LDC) may be purchased on a nonfirm basis. Pricing by the LDC may be based on competitive alternatives similar to interruptible commodity pricing or may be offered at a firm cost-of-service-based price.

- *Load control, or load management, rates*. These rates are a type of IR that gives the utility direct control over specific loads (sometimes via radio wave signals) during system peak periods. This strategy is most commonly used by electric utilities for residential water heating and air-conditioning customers, but can be used for commercial, industrial, and agricultural customers. Load control is intended to minimize customer inconvenience by selecting loads that can most easily be eliminated or cycled.
- *Standby rates*. Utilities offer these rates to customers that have their own source of energy but require service from the utility on an intermittent basis. These rates are more common to electric utilities and may be categorized as maintenance rates for power supplied by the utility when the customer prearranges downtime for generation equipment maintenance; supplementary rates used by self-generators that regularly require additional power from the utility; and backup rates used by self-generators in the event of an unexpected system outage.

NEGOTIATED AND SPECIALTY RATES

Negotiated rates may be cost-based, designed to allow the utility to compete with alternatives, used to support economic development or business recovery, or implemented to permit unique arrangements, such as interruptible service.

Cost-based negotiated rates are designed for customers whose usage and characteristics vary considerably from the average of the rate class or have realistic competitive alternatives to the utility offering the rates. Most commonly, this is only done with large CI&I customers. In many cases, the utility has the ability to negotiate rates down to a level equal to or, more commonly, slightly above its short-term marginal cost. The regulatory justification is that the rest of the utility's customers will benefit from such contracts as long as the negotiated rate charged to the particular customer covers the incremental variable cost of service and provides some contribution to fixed costs. These rates provide the utility with a maximum degree of flexibility to market their product to customers with special needs or competitive alternatives. Sometimes, particularly when the contract period is longer than 5 years, these rates may be designed to recover long-term marginal costs. With these rates, a larger minimum demand charge is required.

- *Economic development rates*. Utilities commonly offer these rates to provide economic incentives for businesses to locate or expand into their home service territory and/or into economically depressed areas. They are often based on a schedule in which rates are initially discounted and then phased into a standard rate over a period of several years.
- *Business retention rates*. These rates are designed to retain customers with either competitive options from other energy sources, self-generation capabilities, or an interest in moving to another state or service territory. Business recovery rates are designed to retain customers in financial difficulty.
- *Conservation and load management rates*. Many utilities offer these rates to customers that meet certain equipment or building envelope thermal efficiency standards and/or operating standards. These rates may be designed with a simple percentage rate break on usage based on achieving a certain level of conservation and efficiency. They may also involve a reduction in charges based on some type of load-control incentive, or they may be based on a combination of both. Rate design also may include incentive mechanisms for shifting load from peak to off-peak periods.
- *Special contract rates*. In cases when it is in the best interest of all parties (i.e., the utility, the customer, and the rest of the utility's customers) and when the unique conditions of the situation cannot be met under standard rules, utilities may develop special contracts with individual customers. Examples are a very large cogeneration application or a customer who makes year-round third-party gas purchases and is willing to

make volumes available to the utility for resale during peak periods. These contracts often require individual PUC approval, which can be a lengthy process.

Examples of other types of specialty rates are compressed natural gas rates for natural gas-fueled vehicles and rates that support the introduction of new technologies.

COMPETITIVE ENERGY RATES

Competitive energy rates give utilities the maximum flexibility to sell power or natural gas in competitive situations. For example, customers considering a gas or steam technology application, such as cogeneration, are often presented with some type of competitive energy rate by their electric utility in an effort to keep the full load on the utility system. Competitive pricing may be offered down to some small level above the incremental avoided cost.

This pricing structure is also used for peak usage by utilities with some excess system capacity. There is some concern that this practice skews market choices, keeps load at the expense of conservation opportunities, discounts opportunity costs, and, in the long run, leads to additional capacity needs at the expense of the rest of the rate payers.

WHOLESALE (OFF-SYSTEM) SALES

Another area for incremental-cost commodity sales is on the wholesale market. This market, while often less stable than on-system sales, can be very profitable for utilities if the plant supplying power or the reserved gas pipeline capacity is in the rate base. A utility with excess capacity in a given period can sell gas or electricity to other utilities or, in some cases, to customers outside of their service territory. The prevailing logic is that excess capacity should be marketed whenever possible as long as variable costs are recovered and some contribution, however small, is made toward fixed cost recovery. With the potential for capacity release, however, LDCs can instead sell excess capacity rights to others rather than use the capacity for the purpose of off-system sales.

REAL TIME PRICING

The RTP is an emerging utility rate strategy that goes a step further than demand charges and varying usage charges in allowing for differentiation of costs that better reflect the utility's actual incremental cost. Real-time pricing rates typically do not have a demand component. Instead, kilowatt-hour, or thousand cubic feet, consumption is priced by the hour. For example, an hourly RTP structure may charge $0.90/kWh at noon on a Wednesday in August and only $0.02/kWh at midnight on a Sunday in March. Currently, RTP is being used by numerous electric utilities.

The theory behind RTP is that if customers are told in advance of the utility's anticipated system and price conditions, customer demand will respond most directly to price changes. That is, a decrease in consumption as the price rises and an increase in consumption as the price falls. Typically, customers are given a schedule of hourly prices one day in advance. In some cases, the cost per kilowatt-hour is fixed for several categories (i.e., off-peak, utility-peak, or regional power pool-peak), but the hours during which they are applied are varied and communicated by the utility to customers on an hourly, daily, or weekly basis.

The procedures used to design rates that differentiate usage and demand charges by time and season of use come close to approximating what the utility determines is proper hourly cost allocation. The RTP accomplishes this with greater certainty and fewer complications. Taking the example of a TOU rate with a peak demand period of the typical 9-to-5 workweek, a peak demand may be set at 9 A.M. by a facility. This peak may have no impact on the utility peak, but is charged as if it were set at the utility system peak hour. With RTP, if in fact this peak had no impact on the utility peak, it would be priced at a far lower level than a peak that did have an actual impact on the utility peak.

The concept behind RTP is that pricing reflects real, almost instantaneous, market conditions instead of predicted market conditions. While TOU rate blocks are bins which approximate what actual costs are in different periods, RTP more closely reflects the actual value of electricity (or gas) at any given point in time. NonRTP rates are, therefore, based on probability of occurrence, rather than occurrence.

Another type of pricing that more closely represents real events is ambient temperature-based pricing. For example, when the outside temperature falls below 30 or 20°F (−1 or −7°C), natural gas pricing could automatically shift to a higher rate. Currently, there are many interruptible gas rates that base interruptions on temperature. A similar strategy could be employed for summer electricity pricing, based on rising temperature. While this is not an instantaneous pricing mechanism, it is one based on real events as opposed to predicted events. When the event occurs, the pricing schedule is in effect.

Electric utility dispatch modeling has become an increasingly precise process. Utilities can identify where each incremental kilowatt-hour comes from and its value. Large facilities can then perform the same modeling of in-house usage. Furthermore, the cost of telemetry is decreasing, while capabilities are increasing. As alternative electricity purchase options become available, RTP may become a mainstream sales pricing tool. Consumers may elect to purchase certain blocks from the utility, generate certain blocks on site, and purchase certain blocks from other sources through retail wheeling, all based on real-time price signals.

RATE RIDERS

In addition to various rate design options, rate riders are special charges or programs integrated into rate schedules that modify the structure based on specific customer qualifications. Riders are used to account for unique conditions or to give the utility added flexibility to apply rates without dozens of additional tariffs. Riders may include: negotiated competitive-energy riders, interruptible riders, standby riders, buy-back riders, conservation and other load-control riders, end-use riders, and other types of discounts, such as an electric utility discount for customers receiving service at a voltage above the standard voltage.

COMMON NATURAL GAS RATE SCHEDULES

Actual natural gas rate schedules include many of the same rate-design strategies previously mentioned. These rates are offered to customers who meet specific criteria. Commonly used natural gas rate schedules are:

- *Firm sales rates*. These are the highest priority of service offered by LDCs. Gas is made available throughout the year, on an uninterrupted basis, to serve customers' needs. Typically, there are several categories of firm sales rates, such as residential, small commercial, and large CI&I. Rate design may include block rates, seasonally differentiated rates, demand charges, etc.
- *Firm transportation rates*. These provide uninterrupted transportation service, through the LDCs distribution system, of natural gas purchased directly by the customer. The LDC takes on the obligation to deliver to the customers' facilities gas that has been delivered to the LDCs gate station by an interstate pipeline.
- *Dual-fuel firm rates*. These provide service to customers who have the option of using an alternative fuel source, but who, at any time, may request firm delivery of gas from the LDC. Since the LDC must stand ready to serve, and may have significant investment in, distribution facilities, supply contracts, or storage capacity, it may require purchase of some level of guaranteed volume or include a demand-charge component. Because this type of service has the potential to negatively affect the LDCs load factor if the customer only uses gas services during peak periods, the rate may be expensive.
- *Interruptible sales rates*. A utility can sell to its interruptible sales customers spot market gas or excess gas that it purchased as a reserve for its firm sales service customers. Selling gas on a commodity basis only allows gas to be priced competitively with oil, propane, and other energy sources. Therefore, customers benefit from lower costs.

  Prices are often negotiated competitively and indexed each month to the alternative fuel or energy source (e.g., No. 6 oil, No. 4 oil, No. 2 oil, propane, or electricity). A rate offering, commonly known as the standard offer, may be made to the entire group with the same energy-source alternative. Individual customers with better purchasing capabilities may negotiate price and be permitted to lock in a rate for a longer or shorter period of time. Prices may also vary with notice period. The shorter the notice period the lower the gas cost.

  The LDC may have the ability to negotiate downward to a certain floor. In many cases, the floor is set a few cents above the LDCs actual supply cost. This competitive approach is accepted by PUCs, because it holds down overall rates by maintaining gas sales that would otherwise be lost to alternative fuels. Rate-case proceedings may include an agreement that a large portion of the marginal revenues from these sales flow back and reduce the rates of firm-rate customers, often via the PGA. In a sense, this flow-back of revenues provides compensation for the use of facilities (fixed costs) that is amortized through cost recovery via firm rate charges.
- *Interruptible transportation rates*. These services are similar to interruptible sales rates, except they relate solely to distribution rather than to both sales and distribution. The LDCs ability and/or need to interrupt are tied to local distribution constraints, not to supply constraints.
- *Cogeneration, air conditioning, and other end-use rates*. These services are offered for specific equipment applications and, because they have a predictable effect on the LDCs load factor and cost, they are typically grouped under individual rate schedules. These rates are designed to be attractive to customers, because they are more closely tailored to actual energy use profiles of the applications, often improving LDC load factor.

COMMON ELECTRIC RATE SCHEDULES

Electric rates use a mix of various rate components, and rate schedules include one or more of the rate design strategies discussed earlier. These components typically consist of an energy charge, a demand charge, a ratchet clause, a FAC, surcharges for factors such as conservation or nuclear plant decommissioning, power factor (PF) charges, and taxes. Often, rates are offered to customers who meet specific criteria, such as type of facility or type of equipment used. Some of the most common non-residential electric rate schedules include:

- *General service (GS) rates*. General service rates are typically used by most small commercial customers. Rate design may include block rates, seasonally differentiated rates, and demand charges. Generally, these rates place a greater emphasis on usage than on demand and are less differentiated than large customer power rates. In some cases, these rates are available without demand metering, using higher usage charges instead. Typically, the availability of GS rates is limited to customers whose demand does not exceed a particular specified level.
- *General service TOU (GST) rates*. General service TOU rates are generally used by small C&I customers with multiple shift operations. They commonly consist of peak and off-peak periods and often register peak demand only during the peak period. They are not time-differentiated as much as are large customer power TOU rates, but they do allow customers to benefit from lower costs for extended use in off-peak periods. They are also often used by customers with electric heat or some type of thermal storage. Rate design may include block rates and seasonally differentiated rates.
- *General service heating (GSH) rates*. General service heating rates are common end-use-specific rates. Typically, they are GS rates that are available only to electric heating customers whose heating-related usage makes up a certain minimum portion of total usage. They are often used by summer-peaking utilities to build winter load. They have a greater degree of seasonal differentiation than standard GS rates, with depressed winter rates compensating for extensive usage. They also may have a TOU component to allow for the use of domestic hot water or heating thermal storage.
- *Street lighting (SL) rates*. Street lighting rates are end-use-specific, typically offered to states, cities, or other municipalities, and sometimes to large campus-type facilities.
- *Large power (LP) rates*. Large power rates are the traditional rates offered to larger CI&I customers and virtually always include monthly or fixed-contract demand charges and may use rate blocks and seasonal differentiation. Typically, the design is somewhat similar to GS rates, except that there is usually increased emphasis on demand charges.
- *Large power TOU (LPT) rates*. These rates are becoming more predominant for large CI&I customers. Typically, they consist of two, three, or four rate periods, such as peak, shoulder, or off-peak. They may register demand only during peak or all rate periods or have fixed-contract demand charges. Off-peak usage may be handled with varied charges or as peak usage with charges in the off-peak periods only for demand in excess of peak demand. Usage charges are varied by rate period. Rate design may include block rates and seasonally differentiated rates.

  These rates are more stratified than traditional LP rates. They are advantageous for facilities with high-load factors and extended hours of operation, which can offset costly peak usage with inexpensive off-peak usage. They are also attractive for facilities that use thermal storage or some type of peak-shaving technology.
- *Real-time pricing rates*. Typically, RTP rates do not have a demand component. Instead, kilowatt-hour may be priced by the individual hour or, in some cases, charges may be fixed, but the hours in which different charges are applied will vary. In either case, the utility communicates these varying costs or hours of application on an hourly, daily, or weekly basis to customers. Often, these rates are used by facilities with alternative energy sources in place or with the ability to shed loads on a regular basis. The RTP rates may become increasingly common as electric rates become even more sensitive to market competition. As opposed to demand-based TOU rates, RTP rates are thought to more closely reflect the actual discrete price of power at a given hour or even minute.
- *Transmission (T) rates*. Currently, T rates are typically used for special cases, such as wheeling, in which power is either bought from or sold to a party other than the local utility. Rate design is based on the use of the utility's transmission facilities only. Under the National Energy Policy Act of 1992 (EPAct 92), electric utilities are required to more clearly define transmission rates, as well as identify available capacity and known restraints. Over the long term, it is anticipated that the advent of retail wheeling and the further unbundling of electric rates will result in transmission/distribution rates being available to all customer classes.
- *Interruptible rates*. These rates and rate riders are designed for customers that have blocks of load (or all of their loads) that can be dropped at any (or almost any) time. Commonly, this includes the use of standby generation as a load-shedding technology. Interruptible rates may be designed with a set rate break on demand or usage, or may consist of a flat annual or monthly fee paid by the utility to the customer with a specific payment rate for each period of interruption. Rate design includes different steps, or levels, of availability, with 100% availability receiving the most beneficial treatment. Rates also vary with notice period. The shorter the notice needed for an interruption, the higher the rate discount.

Further refinement of IR involves differentiation of services between nonfirm electricity sales and nonfirm transmission/distribution services. Facilities with on-site energy alternatives can benefit from the ability to withstand sales and transmission interruptions. Facilities with alternative electricity purchase options, via retail wheeling, can benefit from the ability to withstand sales interruptions, but may still require firm transmission/distribution services.

- *QF rates and rate riders*. Many utilities have special QF rates or rate riders for self-generators. In some cases, these are elective rates (or riders), while in other cases, they are required. These riders often include rate designs that emphasize high peak demand charges, such as TOU rates. These special QF rates also usually include mandatory provisions that require the self-generator to purchase a form of backup services to provide a payment stream to the host utility for providing a reliability service by virtue of the physical connection. The provision of backup service is generally desired by the customer to prevent interruptions. The charge for backup is supported by the argument that a rate recovery mechanism is needed to prevent self-generating facilities. from taking advantage of utility capacity supported by other rate-paying classes, in the event of an outage of self-generation equipment, or during periods of planned system maintenance or for purchasing supplementary power.

  Rate restrictions are often put in place to limit rate options available to self-generators. A self-generator forced to be on a highly demand-sensitive rate with a ratchet penalty clause could end up with nearly a full year's worth of demand charges for a single outage. This is often considered excessive recovery. On the other hand, a self-generator allowed to be on a low-demand, highly usage-sensitive rate may pay a minimal amount, which is often considered insufficient cost recovery.
- *Standby rates*. These rates are offered to self-generators requiring power from the utility when their own or alternative energy supply is inadequate or unavailable. Standby rates are often riders which affect several rates offered by a utility. Many utilities have special rate recovery treatment for providing standby (QF backup) power to self-generators when their own or alternative supply is inadequate or unavailable. Standby charges are a type of demand (or insurance) charge paid to the utility to reserve replacement capacity and energy if a system failure or normal maintenance interval takes the on-site generator out of service. These standby rates may be offered as separate rates or rate riders.

There are three general types of standby rates offered by electric utilities: backup, maintenance, and supplementary rates. Backup and maintenance rates are offered to provide power when a self-generator's system is fully or partially out of service. Maintenance rates are offered for use during prearranged downtime and backup rates are offered for unanticipated downtime. Supplementary rates offer power for regular use and are offered to partial-requirements customers that may require purchased power in addition to their own self-generated power.

Standby rates for backup and maintenance are usually based on a monthly charge per kilowatt of capacity reserved. There is also a commodity charge for actual energy usage during the down-time period. These standby demand charges are less costly per kilowatt than the actual demand charges on a given full service rate, but are paid for on a take-or-pay basis, regardless of whether additional power is ever required.

Maintenance service rates provide convenience in that they allow self-generators to perform routine service and overhaul during peak demand setting periods. An alternative is to perform maintenance in off-peak periods. However, this is not always possible, particularly for lengthy overhauls. Backup service rates are somewhat like an insurance policy. By purchasing this capacity insurance for a given amount of kilowatt on a monthly basis, self-generators avoid ratcheting and/or full demand charges that might otherwise result from outages.

Consider an example in which the full service rate demand charge is $18/kW/month and the standby charge is $8/kW. The facility pays this charge regardless of whether backup power is used. In this example, the annual fee of $96/kW would be a wise investment only if the facility experienced peak setting outages more than 5 month/year, since 5 months' demand charges would only cost $90/kW.

The cost of standby service varies widely. Some utilities require self-generators to purchase standby insurance. Other utilities offer it as an option, while some have no provision for standby power at all. The alternative is the use of standard rates. Key questions in cost allocation are: "What are true costs?" and "What is a fair and reasonable price for such capacity insurance?"

The logic behind this particular cost allocation is that the utility must stand ready to serve these loads when needed. Cost allocation is based on a determination of the impact on generation and distribution capacity requirements. The cost-of-service analyses take into account all self-generators and the real probabilities of peak demand impact resulting from random system outages. If, for example, each of 100 self-generators were to set a peak once a year at different times, what would the real impact be on the capacity requirements of the utility?

For cases in which standby service is not mandatory, but offered as an option, customers must make the determination whether to take this type of insurance or take their chances on standard rates. Customers may also elect to secure standby power for a portion, rather than all, of their self-generation load.

In cases where standby charges are mandatory and very high, the cost may be sufficient to make projects uneconomical. In some cases, a change to mandatory requirements, resulting from rate case proceedings years after a system has been installed, may provide sufficient incentive to abandon a project due to the evaporation of savings critical to successful economic operation.

One hypothetical example of such prohibitive effects is a system with three generation units with required standby charges for the full connected load at 66% of the standard demand charge. In this case, the cost of standby service is equivalent to that of the system operating on a standard rate and experiencing the highly unlikely occurrence of a peak-setting outage in every single month for two out of the three units. Add to this, the potential of being forced onto an uneconomical rate, and a self-generator could end up with no savings at all. While this example is extreme, it helps to explain why self-generation has been underdeveloped in certain utility service territories.

Real-time pricing rate structures may offer an effective means of allocating costs for standby power. Real-time pricing is an attempt to reflect short-term costs so that consumers may make short-term purchase decisions. These same varying short-term prices could be made available to QFs. California, Florida, and Virginia are a few of the states that currently have QF purchase power pricing tied to variants of RTP. For example, one rate structure on file with a state commission provides payment for a QFs energy sales at the corresponding marginal cost (i.e., system lambda, $/MWh) of power the host utility experiences. In California, a forecast of marginal costs are the primary input in determining an RTP pricing structure for as-available energy.

MEASURING ELECTRIC DEMAND

The measurement of demand is fundamental to most electric rate structures. It is a tool that allows a utility to differentiate capital cost requirements for serving customers of varying usage patterns. By measuring and billing for demand, the utility can assign costs more fairly to customers.

As opposed to gas utilities, the demand interval for electric utilities is extremely short, usually 15 or 30 min. It is not an instantaneous measurement, but an average of discrete measurements over time. If, for example, a facility experienced a rate-of-use pattern of 200 kW for 5 min, 300 kW for 5 min, and 700 kW for 5 min, a 15-min demand interval meter would register an average demand of 400 kW. The demand recording meter would log 400 kW and reset only when a higher level of demand was reached. Sometimes, utilities set peak demand by averaging peaks of a few demand intervals over the billing period.

Many customers use demand monitoring and load-shedding techniques to minimize the impact of peak demand billing. These facilities often attempt to synchronize their operations with the utility demand interval and use intermittent load shedding to reduce their average rate-of-use during intervals when a large surge of power is required for a period less than the full demand interval.

From the utility perspective, this load shedding technique may partially defeat the purpose behind demand metering, which is to charge for peak capacity requirements. A sliding demand interval is sometimes used to more accurately measure the impact of peak demand. With a sliding demand interval, for example, a 15-min demand interval may be broken into smaller intervals of 5 min. These smaller intervals are averaged and then added together, as in the previous example, to set the peak demand for the entire 15-min interval.

In some cases, utilities simply use smaller demand intervals, starting as low as 5 min. More common, however, is the use of the typical 15- or 30-min interval with a clause in the rate schedule that states that in cases of rapidly fluctuating loads or other special conditions in which the established demand measurement time interval does not equitably compensate the utility, demand may be based on the peak for a shorter period.

Traditional rates often use only one peak demand measurement for billing. Some rates call for demand measurement only in certain peak periods. The rationale is that individual facility peaks in utility off-peak periods have no real impact on capacity requirements. The TOU rates, however, measure peak demand in several periods. Demand charges may be set at a different rate for each period. For example, peak demand might be billed at $20/kW during peak periods and $5/kW during off-peak periods. In some cases, off-peak period peak demand may only be billed for the portion that exceeds peak period peak demand. This is referred to as excess demand billing.

INTEGRATING POWER FACTOR INTO DEMAND BILLING

Utility generation is measured in volt-amperes, or apparent power, while most customer meters are measured in watts, or real power. In alternating current (a.c.) circuits, watts (power) are equal to volts (potential) times amps (current) only when the wave-forms of voltage and amperage are in phase. This is an ideal condition that does not exist in electric distribution systems. Many types of equipment, such as induction motors, require more apparent power than the amount of real power consumed because their inductive impedance causes current and voltage to be out of phase.

The difference between apparent power volt-amperes and real power watts is called volt-amperes reactive. This is the component of volt-amperes that circulates back and forth between the utility and the equipment, but is not consumed by the load. It is, however, partially consumed by distribution losses.

The PF is the ratio of W:VA. A facility with a PF of 0.75, requiring the same wattage as another facility with a PF of 1.0, for example, will be more costly to serve, because the utility will require one-third more system capacity (1.00 W/0.75 PF = 1.33 VA) to serve the facility with the PF of 0.75.

To more accurately allocate capacity costs through demand charges, some utilities measure and bill demand charges based on kilovolt-amperes rather than kilowatt. This shifts the cost for maintaining nonproductive capacity, or a PF of less than 1.0, to the customer and acts as an added incentive to improve the facility's PE.

Many utilities simply institute a penalty for a lagging PE For every increment below a required minimum PF, a charge is levied against the facility. The minimum allowable PF is typically in the range of 0.80–0.90. Many utilities establish this penalty on the rate schedule, but often do not invoke it.

Another way utilities establish a PF penalty is to specify a maximum free kilovolt-amperes reactive as a percentage of the maximum kilowatt of demand. Utilities then bill for all metered kilovolt-amperes reactive above this level. There are several other ways to build PF into billing, such as increasing the peak demand measured by a certain percentage for every percentage, the PF is below a specified level.

Many utilities currently do not penalize for lagging PF, and many that do only impose modest penalties. In those cases, from the customer's perspective, the advantages of a higher PF and the benefits of investment in capacitors and other higher PF equipment are savings from reduced internal line losses, down-sized equipment, and avoided billing penalties. To encourage such customer actions, utilities attempt to set PF penalties at levels that will have sufficient economic impact. Refer to Chapter 24 for a detailed technical description of PE.

METERING POINT AND TRANSFORMER OWNERSHIP

Another factor that is often an element in electric utility rate structures is metering point and transformer ownership. Utility distribution voltage is almost always greater than the voltage required inside a facility. The main transformer brings the voltage down to a suitable level for the service entrance at the facility. Typically, the utility owns the transformer and meters usage on the customer (low-voltage or secondary) side of the meter.

These factors will figure into rate design. If the customer owns the transformer, the utility saves on capital and maintenance costs and can pass those savings on to the customer. If power is metered on the high side of the transformer, the utility saves on transformer-related power losses and can pass those savings on to the customer to compensate for losses now occurring on the facility side of the meter.

SEPARATION OF COMMODITY AND DISTRIBUTION FUNCTION

Overtime, the forces of deregulation are moving the jurisdiction of regulated utility rate structures toward the transmission and distribution functions and away from the commodity sale function, which is falling under the control of free market forces. However, in today's market, the rate structures presented above still widely apply as the utilities continue to sell a bundled commodity to their customers or sell unbundled services, notably transmission and distribution under regulated cost-of-service-based pricing.

With the advent of electricity wheeling and gas brokering, with open access to transmission and distribution service, opportunities for utilities to make off-system sales have become subject to fierce competition. With open access for the transmission of power for wholesale sale in interstate commerce, the utility must compete for wholesale market share with cogenerators and other independent power producers, many of which can generate power at a very low cost. Gas brokers present similar challenges to gas utilities.

From the consumer perspective, the trend of competitive pricing can be very attractive. Currently, competitive pricing is most beneficial to large consumers, notably those with favorable load profiles or competitive options. Such customers often face a win–win situation in which they install operating cost-reducing alternatives or reap savings from a competitive energy rate break or, in some cases, both.

Over time, it is anticipated that an increasing number of consumers will be the beneficiaries of competitive pricing. However, there is also concern that the embedded cost in utility capacity that is no longer cost-competitive will drive up prices to customers stranded without strong competitive options.

The overwhelming long-term trend in the electric and the gas industries is one of deregulation, fierce market competition, unbundling of services, and increased consumer choice in energy services. Historically, natural gas companies mostly sold and delivered purchased gas to customers at a price determined by a regulated rate structure that included a bundled host of services. Electric utilities sold electricity that they had generated or purchased under similar arrangements.

While far from fully evolved, the current trend is toward a condition under which the local utility's prime function is the transmission and/or distribution of natural gas or electricity, but not necessarily the sale of the energy commodity. The purchase and sale of natural gas and electricity is becoming more of an independent market function in which the utilities are but one broker among many. It is very conceivable that over the next decade, this trend will be all encompassing, inclusive of the residential market, with newly developed rate structures that are compatible with the evolving conditions surrounding consumer transactions. Though the trend continues at an accelerated rate, utilities still provide both sales and distribution functions to most customers.

CONCLUSION

The concepts behind the design of natural gas and electric utility rate structures are extremely similar in that they are developed through the same regulatory process. Both have been greatly affected by the movement toward deregulation and are moving more toward market-based competitive rate structures. While the pace of change varies from state to state and utility to utility, the commodity component is being separated from the transmission and distribution component, necessitating significant changes in cost recovery and, therefore, rate design.

Finally, the trends toward some variation on the concept of RTP are apparent, as is the trend toward increased customer choice in the selection of utility rates and energy services in general. Still, the basic concepts of fixed and variable cost recovery shall continue to apply. An understanding of these concepts allows gas and electricity consumers to understand utility rates and select rates and services that most effectively meet their energy-use needs.

BIBLIOGRAPHY

1. Berg, S.V. *Innovative Electric Rates*, Lexington Books, D C Heath and Company: Lexington, MA, 1983.
2. Berg, S.V.; Tschirhart, J. *Natural Monopoly Regulation: Principles and Practices*; Cambridge University Press: New York, NY, 1988.
3. Bonbright, J.C.; Danelson, A.L.; Kamerschen, D.B. *Principles of Public Utility Rates*, Public Utility Reports, Inc.: Arlington, VA, 1988.
4. Capehart, B.L.; Turner, W.C.; Kennedy, W.J. *Guide to Energy Management*, 5th Ed.; The Fairmont Press, Inc.: Lilburn, GA, 2006; [Chapter Three, Understanding Energy Bills].
5. Enholm, G.B.; Malico, J.R. *Electric Utilities Moving Into the 21st Century*, Public Utility Reports: Arlington, VA, 1994.
6. Hirsh, R.F. *Power Loss—The Origins of Deregulation & Restructuring in the American Electric Utility Stystem*; MIT Press: Cambridge, 2002.
7. Petchers, N. *Combined Heating, Cooling & Power Handbook*; Marcel Dekker, Inc.: New York, NY, 2003; [Chapter 21, Utility Rate Structures, and Chapter 22, Utility Bill Analysis].
8. Turner, W.C. *Energy Management Handbook*, 5th Ed.; The Fairmont Press, Inc.: Lilburn, GA, 2005; [Chapter 18, Electric and Gas Utility Rates for Commercial and Industrial Customers].
9. The Changing Structure of the Electric Power Industry 2000: An Update; October 2000, US Energy Information Agency, Washington, DC.
10. Introduction to Rate Design for Electric Utilities, Professional Development Course, Electric Utility Consultants, Inc. Denver, Co., 2006.

Walls and Windows

Therese Stovall
Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.

Abstract
Energy travels in and out of a building through the walls and windows by means of conduction, convection, and radiation. The walls and windows, complex systems in themselves, are part of the overall building system. A wall system is composed of multiple layers that work in concert to provide shelter from the exterior weather. Wall systems vary in the degree to which they provide thermal resistance, moisture resistance, durability, and thermal storage. High-tech windows are now available, which can resist radiation heat transfer while still providing light and visibility. The combination of walls and windows within the building system can be adapted to meet a wide range of environmental conditions, recognizing that the best building envelope system for one climate may not be the first choice for another location.

INTRODUCTION

The building envelope protects you from the weather, separating the indoor air that you have paid to heat or cool from the outdoor air. The walls, roofs, and windows form the major surfaces of this envelope. Energy travels through these surfaces along many paths and in many forms, such as unintended air leakage through a wall or sunlight streaming through a window. We can improve the energy efficiency of the building envelope if we carefully consider all of these energy pathways.

Here, we are focusing on energy efficiency, but keep in mind that each part of the building envelope performs multiple jobs under challenging conditions. The roof keeps out rain, hail, and snow; bakes in the hot summer sun; freezes in the cold winter night; and must be sturdy enough to survive a workman's boots. The walls must repel the rain, hold up the roof, stop the wind, and provide a rigid support for windows and doors. The windows have to let in the light, allow ventilation when they are open, and keep out drafts when they are shut. Any change we make to the building envelope to conserve energy must account for these multiple functions and the complex interactions between the envelope components.

Resistance to heat transfer is often expressed as an *R*-value. For a complete description of R-values and their use, please refer to the DOE Insulation Fact Sheet (www.ornl.gov/roofs+walls/insulation).

BUILDING TYPES

Buildings fall into two main classes: high-rise and low-rise. The high-rise buildings are typically custom engineered with structural steel frames. Low-rise buildings are typically divided into commercial, low-rise multi-family, and single-family buildings. The high-rise, commercial, and low-rise multifamily buildings are more likely to have low-slope (often mislabeled as "flat") roofs, whereas the single-family houses are more likely to have pitched or steep-sloped roofs.

Construction methods can be roughly grouped according to the portion of the assembly performed on site and the portion performed at a factory. With today's engineered wood products and premade trusses, few buildings are strictly built on site; but we still refer to a stick-built building as the one where the greatest part of the assembly takes part on the construction site. At the other end of the spectrum are manufactured buildings that can be moved from one site to another with relatively little effort. In the middle are factory-built modular buildings and panelized construction. A factory-built modular building typically includes one or more modules that are placed on a permanent foundation. In some modular buildings, the windows are installed after the modules have been installed. A panelized building is closer to the site-built model, but will have major wall, floor, or foundation sections prebuilt and delivered to the site.

Every building must conform to local building codes, which can limit the material choices or construction methods. Many local building codes now include energy conservation clauses or incorporate the Model Energy Code or the International Energy Conservation Code.[1]

WALLS

The walls make up most of the exterior surface area of many buildings and therefore the energy transported through these surfaces is very important. Wall issues vary according to the building type. A framed building is constructed with a wood or metal skeleton that provides structural support to both the building and all the other wall components.

Keywords: Building envelope; Wall; Window; Thermal energy; Heat transport; Solar heat gain coefficient; Thermal mass; Adobe; Structural insulated panel; Exterior insulation finish system; Masonry walls.

A framed wall is characterized by numerous parallel heat paths and multiple layers of different materials. A non-framed building uses bulk material, such as masonry or adobe, to provide the structural support. This type of building is characterized by a more homogenous heat path and relatively few layers of materials.

For either type of building, the connections between the wall and the roof and between the wall and the foundation are important construction details from an energy conservation standpoint. These connections can provide unintended air passageways and may be overlooked in the overall insulation scheme. An Air Drywall Approach is an effective way to limit air leakage from walls, roofs, and windows. Here, a rubber gasket is fitted along the perimeter of the window frame and along the exterior wall's base board and ceiling plate to compensate for openings that occur as the wood changes shape with time. The gasket, once placed, seals against the drywall gypsum board and makes an airtight barrier.

Whether the wall is built on a frame or constructed from masonry, there is a wide selection of exterior siding choices. These include brick; wood, fiber cement, or vinyl siding; or an exterior insulation finish system (EIFS). For all of these facades, repelling rain and wind is often more complex than it looks. For example, a brick wall looks like a solid surface, but the mortar joints provide capillary paths for moisture, especially when subjected to wind-driven rain. Similar pathways exist for other cladding materials. For this reason, air gaps are often provided behind the outermost wall layer. Depending on the vent/drain arrangement, this may provide a true pressure-equalized rain screen or a simple break in the capillary path. Old-fashioned wood lap siding is a good example of a simple rain screen, as shown in Fig. 1. The small air gap behind each wood layer is well vented and drained to the outside, so that the air pressure within the air gap is equal to the air pressure outside the wall. This pressure equalization reduces the moisture moving into the air space, and therefore reduces the amount of moisture available to penetrate the rest of the wall.[2,3]

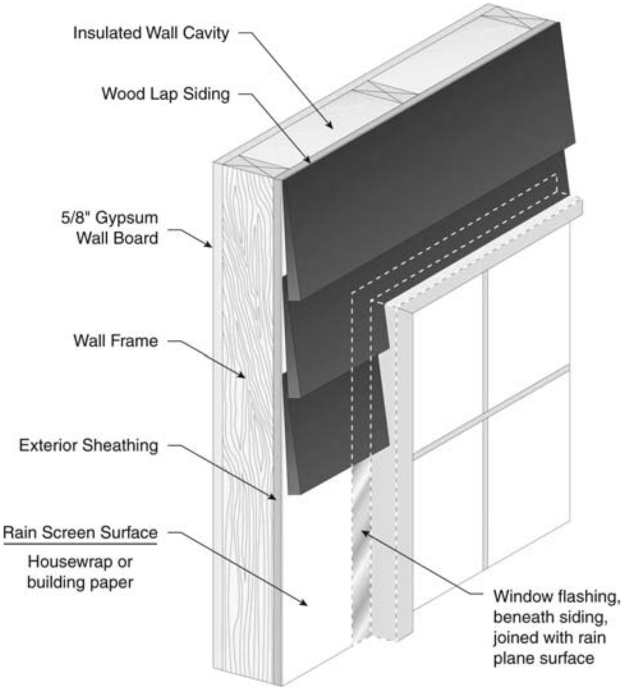


Fig. 1 Simple rain screen within a typical wood lap siding wall.

Proper moisture management is important for energy conservation for two reasons. First, unmanaged moisture must be removed by additional ventilation, which entails the energy load needed to heat and cool the additional air mass. Second, moist building materials will always have a higher thermal conductivity than dry materials. When wet, some insulation materials become matted and lose the greater part of their insulating value.

Every wall system stores energy. This storage quality is called "thermal mass." The thermal mass can reduce the amount of heat lost or gained through the wall whenever the outdoor temperature varies above and below the indoor temperature on a daily basis. This occurs during the spring, summer, and fall for most of the United States, and during the winter as well for the southern regions. The relative benefit of thermal mass is therefore determined by both the wall properties and the climate. A wood-framed wall has very little thermal mass compared with a masonry wall. One study compared the energy consumed by a house with traditional wood-framed walls with a house constructed with masonry walls for six cities. Depending on the location and wall thickness, the more massive wall reduced the household energy use by an amount equal to increasing the traditional wall's thermal resistance by 10–50%, with the savings greatest in Denver and Phoenix and least in Miami.[4]

Although we typically think of the exterior walls when we think about energy losses, the interior walls are also important. Air enters the wall cavity through a number of penetrations, some visible and some not. Holes made in the drywall to accommodate electrical outlets and plumbing connections also allow uncontrolled airflow. Other gaps are often present at the floor–wall and floor–ceiling connections. Therefore, it is important to provide a continuous top plate above every wall cavity. This top plate separates the interior wall cavity from the attic space, thus preventing a free flow of conditioned air from your house into the attic.

In addition to the energy used to heat and cool a building, energy is also embodied in the building materials and expended in the construction process. Among low-rise residential buildings, studies have shown that many of the building elements, including gypsum drywall, roofing materials, and carpeting, are common to all the wall types. But the wall type still makes a significant difference in the overall energy embodied in the building, with a wood-framed house containing about 15% less embodied energy than either a metal-framed or concrete house.[5,6]

Wood-Framed Walls

Wall construction methods and materials vary somewhat according to the local climate and natural resources, but most walls in the United States are made from wood framing with insulation between the studs, drywall on the inner surface, and exterior sheathing layer(s) (Fig. 2). The wall studs used are either nominal 2×4, with a 3.5-in. cavity depth, or nominal 2×6, with a 5.5-in. cavity depth. Cavity insulation options for new construction include batts, a blown-in mixture of a foam binder and loose-fill insulation, blown-in loose-fill insulation secured by nets, or blown-in foam insulation.[8] In commercial buildings, high-density batts are sometimes used to provide both thermal insulation and acoustical buffering. In retrofit situations, professional installers can blow loose-fill insulation into the wall cavities by drilling a series of holes through the interior or exterior facade between each pair of adjacent studs.

Energy flows through the wood studs more easily than through the surrounding insulation. Most walls contain much more framing than you would think, as shown in Fig. 3. In addition to the studs, there is additional wood framing around each window, around each door, at each building corner, where the wall sits on the foundation, where the roof sits on the wall, where an interior wall meets an exterior wall, and between floors in a multistory building. When you combine all of these thermal "short circuits," a wall (in a one-story ranch style house) filled with R11 insulation provides an overall performance of only R9–R10. If you replace that insulation with foam in the wall cavity, thereby increasing the cavity insulation from R11 up to about R18, you get an overall performance of R13 or an increase that is about half of the increased insulation value. One way to improve the thermal performance of a wall is therefore to place insulation between the studs and the exterior surface of the wall. For

Fig. 3 Framing lies behind a significant portion of the wall area in many houses.
Source: From ASHRAE Special Publications (see Ref. 9).

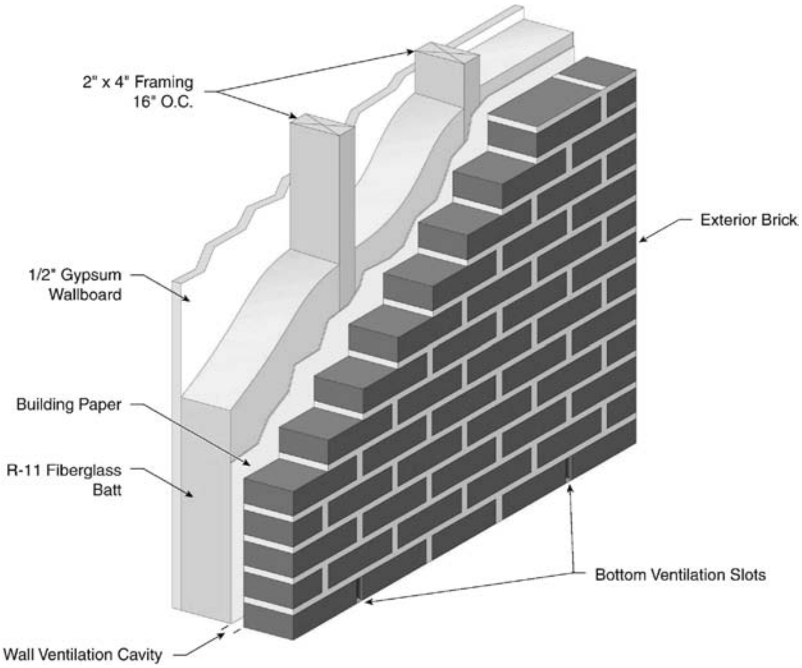


Fig. 2 Typical wall structure for a wood-framed wall with brick cladding.
Source: From ASHRAE Special Publications (see Ref. 7).

Util—Wall

example, adding only 1 in. (R5) of foam sheathing to the R10 wall brings its overall *R*-value up to R14 (Fig. 4).

Advanced framing techniques are available which reduce the amount of lumber required. These methods reduce the energy losses through the framing and allow more room for insulation. Many of these techniques also provide for improved air sealing.[10]

The wall sheathing provides a flat uniform surface to support the exterior air barrier, vapor retarder, and siding. If a wood product is used for the sheathing, it will also provide the structural stiffness needed at the building corners. If a foam insulation product is used instead, some form of additional bracing will be necessary at the corners. Sometimes a layer of foam is placed on top of a layer of wood product sheathing. This greatly improves the wall's thermal resistance because that layer covers the thermal short circuits provided by the wood frame. However, this configuration can require extra care during the finishing process, with longer nails needed to fasten siding materials. Also, specialty brick ties will be needed if a brick decor is selected for finishing the wall.

Steel-Framed Walls

Steel-framed walls share many of the characteristics of a wood-framed wall, but the steel components themselves have a very high thermal conductivity. Therefore, most steel-framed walls are built with a layer of foam sheathing to break that thermal pathway. There is also research underway to produce complex steel shapes to provide the structural support needed while providing a longer heat transfer pathway, and therefore greater thermal resistance.[11] Some metal-framed products also include an integral foam insulation element for the same reason. Connections between the walls and a steel-framed roof can be problematic from an energy point of view, especially if the steel framing extends out in the eaves. Such arrangements act like the fins on a heat exchanger and can

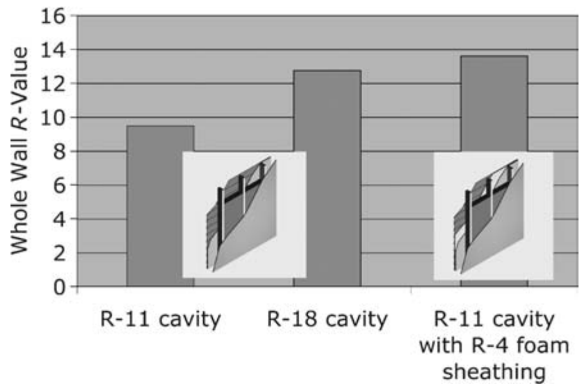


Fig. 4 Whole wall *R*-values for wood-framed (nominal 2×4) walls on 16-in. centers with interior gypsum and exterior wood siding in a one-story ranch house.

cause excessive energy consumption if appropriate insulation arrangements are not included in the design. Steel-framed walls are most attractive where the heating loads are modest and where insect damage is more challenging.

Masonry Walls

Masonry walls can refer to walls with a masonry veneer, such as brick or stone, or to a wall where the masonry also provides the structural support, such as a poured concrete or concrete block wall. Such buildings are relatively resistant to the corrosive environment common near the ocean. Masonry also provides thermal mass that can both save energy and improve the interior comfort level in a hot climate with daily temperature swings. Aside from this thermal mass effect, masonry veneer walls share the same energy characteristics as other wood- or steel-framed walls.

Concrete block and poured concrete

Full masonry walls are often used for foundations or basements. Whole houses built from masonry are popular in the warmer climates where termites and other insects are more populous. A full masonry wall can be built from concrete blocks or by pouring concrete into forms. For a load-bearing wall, steel reinforcement rods, or rebar, are used with either method to add strength. For a reinforced concrete block wall, reinforcing rods are placed vertically in the block cavities which are then filled with mortar. Steel wires or mesh are laid in the horizontal mortar joints to resist shear stress. In a poured concrete wall, the steel reinforcing rods are positioned both vertically and horizontally within the forms before the concrete is poured. Because these steel rods tend to be perpendicular to the heat transfer direction, they have little effect on the overall wall thermal resistance.

The thermal characteristics of a masonry wall vary, depending on the whether blocks or poured concrete is used and on the density of the concrete. But in general, the thermal resistance of the masonry portion of such walls will be very small, ranging from R1 to R2 for an 8-in. thick wall, even if the cores of a hollow concrete block are filled with perlite or vermiculite.[12] Foam board insulation, with a thermal resistance in the range of 5*R*/in. is often used to increase the total thermal resistance of a masonry wall and can be placed on the inside and/or the outside surface. The foam board must be covered by some material with an appropriate fire rating, such as gypsum board. One study has shown that the thermal mass is more effective when the concrete is in good thermal contact with the interior building, i.e., when the insulation is placed on the outside of the wall.[13]

Autoclaved aerated concrete can be used in place of ordinary concrete for low-rise buildings. This type of concrete is much lighter than standard concrete, available in a variety of sizes and shapes, and may be reinforced. The autoclaved aerated concrete has a much higher *R*-value (1.25*R*/in.) than the standard concrete (0.05*R*/in.).[3] The overall *R*-value of a wall built with this material will depend on the shape and thickness of the concrete and on the thermal resistance of other wall components, such as air cavities. Several walls tested with aerated concrete (no facings applied), both in the traditional hollow concrete block and solid block forms, had *R*-values between 6 and 9.[14]

Precast concrete

Walls can be made from reinforced concrete slabs that have been precast into the desired shape. Unless insulation is applied, these walls will have approximately the same thermal resistance as a site-poured concrete wall, i.e., about R1–R2 for an 8-in. thick wall. This construction method is more often used for larger buildings, such as apartment buildings or hospitals, but is also used for residential buildings. In the smaller buildings, the precast panels were first used for basement walls, but precast panels can be used for all exterior walls. The panels are precast and cured in the factory, thus avoiding weather limitations associated with pouring and curing concrete at the building site. The factory environment also allows the production of concrete that is stronger and more water resistant than site-poured concrete. Some of the panels are cast against foam insulation to improve the wall's *R*-value. In addition to the steel reinforcing, these walls can be produced with cavities for electrical wiring and rough openings. Some of the panels have been designed to provide the appearance of bricks and limestone, so that a new building will blend into an existing urban neighborhood. A crane is usually used to place the precast panels on top of a bed of crushed stone. The panels are then connected in place using weld joints or bolts and sealants. Installation for a typical residential unit can be completed in a single day.[15,16]

Insulated concrete forms

An insulated concrete form (ICF) wall is composed of a set of joined polystyrene or polyurethane forms that are filled with reinforced concrete, as shown in Fig. 5. The joining methods vary from one manufacturer to another. Some have interlocking foam panels; others are linked with ties made from polypropylene or steel. The foam panels become a permanent part of the wall, providing a continuous layer of insulation on both the inner and outer wall surfaces. The concrete portion of the wall is reinforced with rebar positioned inside the forms before the concrete is poured to provide needed strength. The ICF walls can be covered on the outside with light-weight stucco, brick, or wood or vinyl siding. Gypsum board fastens onto the interior side. The attachment methods for

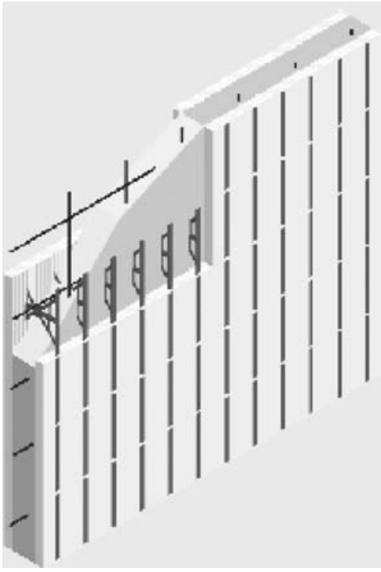


Fig. 5 Insulated concrete form (ICF) wall.

these facing materials vary from one manufacturer to another. In the laboratory, the simple steady-state *R*-value of unfinished ICF walls varied from 12 to 18.[14] The time-varying effective energy performance of these walls is also determined by the temperature profile within the wall, which is improved by the walls' well-insulated thermal mass. Therefore, the ICF wall design provides both greater thermal mass and a higher thermal resistance than a typical 2×4 frame wall. Although the airtightness of any wall system tends to vary depending upon the expertise of the construction crew, an ICF wall is generally more airtight than a wood-framed wall.

A variant of the ICF wall uses forms made from a polystyrene–cement composite. The composite walls may have a greater thermal mass, but provide a significantly lower steady-state thermal resistance (about R8 for one 10-in. thick configuration).[14]

Adobe

In parts of the southwest, adobe walls are popular because they absorb and store the daytime heat until it can be released to the cooler night air. Traditional adobe bricks are sun-dried, not fired, mixtures of clay, sand, gravel, water, and straw or grass. (For fired or stabilized bricks that are made to look like adobe, see the previous section on concrete masonry). The traditional production method produces a brick that swells and shrinks depending on its fluctuating water content. Adobe walls are relatively fragile and must be sealed with a protective covering.

The low compressive strength of the adobe bricks leads to the use of very thick (10–30 in.) walls that are seldom more than two stories high. These thick solid walls provide significant thermal mass. The thermal mass effect is especially important because the steady-state thermal resistance of these walls is not very great. A 14-in. thick wall would have a thermal resistance of from R2 to R10, depending on the density and water content of the wall. Insulation can of course be added to the interior or exterior face of the wall if covered with an appropriate coating.[17]

Exterior Insulation Finish Systems

The EIFS can be placed on a wood- or steel-framed wall or a masonry wall. In this system, a layer of polystyrene board insulation, one or more inches thick, is applied to the wall which is then covered with multiple coatings that produce the finished appearance of stucco. This wall system provides a continuous cover of insulation that breaks all the thermal short circuits associated with the framing materials and has the thermal advantage previously shown in Fig. 4 for foam sheathing. The EIFS system has been and continues to be one of the most popular exterior claddings for commercial and institutional buildings. However, residential construction jobs often do not have the same high level of quality control and job oversight. This difference led to moisture-related problems in residential EIFS walls where moisture seeped into inadequately sealed window openings and became trapped within the walls.

Subsequent building failures led to the development of two classes of EIFS: barrier and drainable. Both classes are used in the commercial building class, but only the drainable system is allowed for residential construction in many locations. In the barrier class, the outer finish layer is designed to be the one and only weather-resistive barrier on the wall. In the drainable class, some form of spacer is placed between the polystyrene board insulation and a second weather-resistive barrier is located atop the wall's structural sheathing layer. This drain plane allows any moisture that seeps into the wall to drain safely out of the wall.[3]

Structural Insulated Panels (SIPS)

The SIPS walls are made by sandwiching foam insulation, typically 4–6 in. thick, between two sheets of a wood product, thus providing structural support, insulation, and exterior sheathing in a single panel. Each manufacturer specifies the proper method and materials to use when joining adjacent panels. Because the panels themselves have such a high thermal resistance, these joints are critical in maintaining a high thermal resistance for the whole wall. Walls have been measured with *R*-values of about R14 for a wall with a 3.5-in. thick foam core and about R22 for a wall with a 5.5-in. thick foam core.[14] The wall sections are relatively light and the exterior walls of a building can often be completed in a day. Two variations of the system include the use of metal sheets for the exterior skin and the use of alternative insulation materials.

Straw Bale Walls

Exterior walls can be made from stacked straw bales. The straw is a natural insulation material, but must be protected from the weather by an exterior surface, often a stucco-type finish applied over a wire mesh. Gypsum board can be used on the interior surface using a number of methods.[3] Experiments have shown that it is very important not to leave any air space between the straw and the surfacing material. Such air gaps work in concert with the hollow straw tubes to set up convection loops that may cut the overall thermal resistance from around R50 down to around R16.[18]

WINDOWS

Transparent glass was first used for windows during Roman times. Technology has gradually advanced, improving the smoothness, strength, and clarity, and increasing the maximum size of manufactured glass. More recently, double-pane windows became popular as energy costs rose and the use of air conditioning became more common. By the mid-1990s, nearly 90% of all residential windows sold had two or more layers of glass.[19] The windows in older homes are being upgraded to double-pane windows as well; from 2000 to 2001, about nine million homeowners spent $15 billion on window and door replacements.[20] From 1990 to 2003, the replacement window market made up from 38% to 52% of the total window market.[21] However, in 2004, two-thirds of residential buildings and about half of the non-residential building stock still had single-pane windows. The potential energy savings in this population is great, especially with the new selective coating techniques for glass.

Energy Transport

The primary function of windows is to let light and air into a building; so it is not surprising that the windows can be very challenging from an energy conservation point of view. Energy travels through a window by all three energy transfer phenomena, as shown in Fig. 6.

The radiation portion of the energy transport includes short-wave radiation, or ultraviolet, that tends to fade the colors in fabric and paint; visible radiation, or light, that we desire; and long-wave radiation or heat. Long-wave radiation is a normal part of the solar spectrum. It helps to heat our buildings during the winter but increases our air conditioning load during the summer. Some values for the solar heat gain through a few prototypical windows are shown in Fig. 7. Long-wave radiation also occurs between any two surfaces, traveling from the warmer surface to the cooler surface. Because the outside environment is warmer than the inside surfaces in the summer and the reverse in

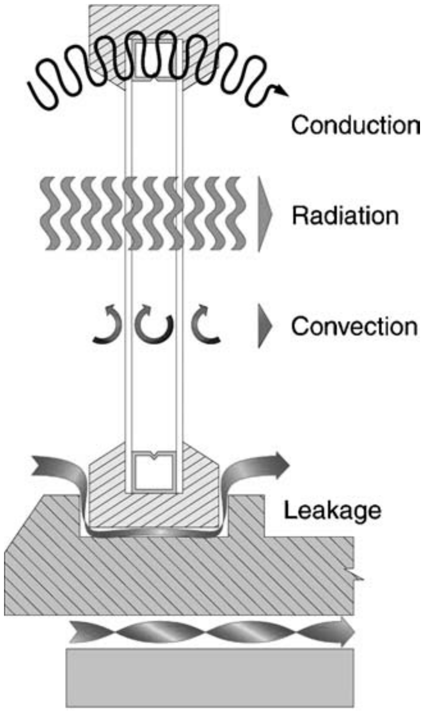


Fig. 6 Energy traveling through windows.

the winter, this long-wave radiation travels through the window and increases both our winter heating and summer cooling energy use. Newer windows have a "low-e" coating to reduce long-wave radiation and thus reduce the energy losses.

The conduction portion of the energy transport includes the heat that travels through the window frame and heat conducted through the glass pane(s) and through any gas between the panes. The energy that travels through the window frame and sashes is a complex function of the frame material(s), the shape (including any hollow cavities), and the exposed surface area. Each material used in windows has positive and negative qualities. In general, wood has a lower thermal conductivity than metal or plastic. However, wood is more likely to change in shape during its lifetime, so that sealing air leakage out of a wood window over a long time period may be more problematic. Also, wood is more susceptible to moisture damage and must be kept painted or varnished. Metal frames will conduct more heat than wood, but the exposed area can be reduced because the metal is a stronger material. Some of the best performance is found in windows that combine multiple materials. For example, a metal- or vinyl-clad window frame will require less maintenance than a wood window, but will conduct less heat than a solid metal or solid plastic frame. Some special gases, such as argon, have a lower thermal conductivity than air and are sometimes used to fill the gap between panes in multiple-pane windows.

The convection portion of the energy transport includes exterior air movement, or wind, across the glass surface; gas movement between the panes in a multilayer window; and air movement across the interior face of the window. The gas movement within a multilayer window is determined by the thickness of the gap, the height of the window, the gas temperature, and the temperature difference between the two panes of glass. Closed draperies or shades can reduce the air flow across the inside pane of glass.

The energy carried by air leakage falls into two major categories. As Fig. 6 shows, some air leaks around the window frame itself. This leakage path should be sealed when the window is installed, although caulking around the frame of an existing window can also reduce the air leakage. Air also leaks through any moving joint in a window. These joints must be sealed by weather stripping or special gaskets that are built into the windows themselves.

Storm windows have been used for a long time and were very popular in northern climates before the introduction of multiple-pane windows. Storm windows require annual installation and storage and there are more glass surfaces to clean. Storm windows can be mounted on either the inside or outside of the window

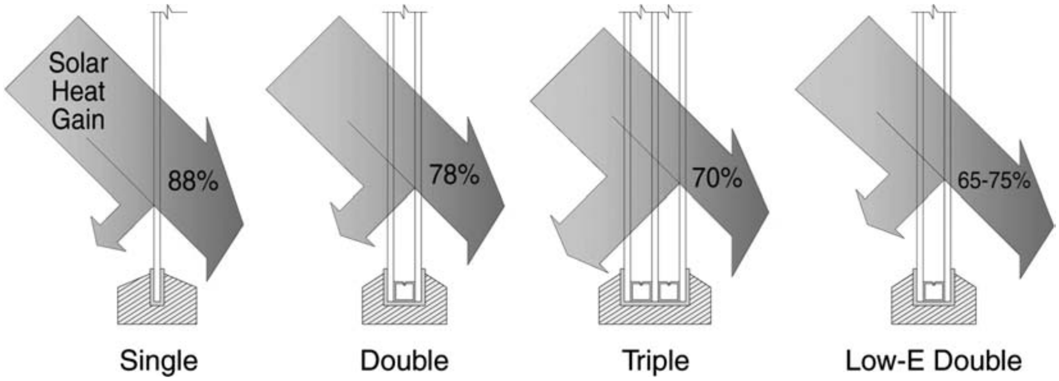


Fig. 7 Solar heat gain through various windows.

frame.[22] Tests have shown that the energy savings can be substantial when storm windows are added to an existing single-pane window. However, replacing an existing single-pane window with a modern double-pane window will save more energy than the addition of a storm window.[23,24]

Window Rating Systems

Considering the complexities of energy transport through windows, it can be difficult for consumers to compare one window with another. Fortunately, there are two important tools available to help with window selection. The National Fenestration Rating Council (NFRC) has developed a rating system that includes a standard label.[25] The U.S. Departments of Environmental Protection and Energy have cooperated in the production of an Energy Star label for windows.[26]

The NFRC label (Fig. 8) shows four values: the *U*-factor, the solar heat gain coefficient (SHGC), the visible transmittance, and air leakage. Manufacturers may also choose to show a value for condensation resistance. The two most important factors used to rate the energy efficiency of a window are the *U*-factor and the solar heat gain coefficient. The *U*-factor is the inverse of the *R*-value (the label that is quoted for insulation), so that a lower *U*-factor indicates a slower rate of heat transfer for any given temperature difference.[27] The NFRC *U*-factor ratings for windows sold nowadays range from 0.2 to 1.2.[28] The SHGC ranges from 0 to 1 and measures how much heat from the sunlight incident upon a window will enter the building. Windows with lower SGHC ratings do a good job of blocking this heat.

The Energy Star label is available to windows that have been certified and labeled by NFRC and that meet special standards. The standards vary among the four NFRC climatic regions because of the trade-off between desirable winter heating and undesirable summer heating. The four regions used by the Energy Star program for windows are shown in Fig. 9 and Table 1 shows the required *U*-factors and SHGCs for each region.[29]

Fig. 8 National Fenestration Rating Council (NFRC) label.

Future Improvements

Researchers are working on a portfolio of window designs with automatic energy-saving features. Some forms of "smart" windows with switchable glazing have become commercially available, albeit at a relatively high cost. These windows vary the amount of light and heat transmitted based upon an electric current, which is usually programmed to respond to either the temperature or the amount of sunlight hitting the window. The electrochromic window uses a multilayer electrically conductive film where ions are moved from one layer to another by a short electrical signal. In one layer, the ions allow only 5% of the sunlight through the window; when the ions move to the other layer, 80% of the sunlight is transmitted. This system has the advantage that once the change from one state to another has been made, no electrical energy is required to maintain that state. Another switchable glazing, the suspended particle display (SPD) places a solution containing suspended particles between two glass panes. When an electrical charge is applied, the particles align and light is transmitted through the window. Without an electrical charge, the particles move about randomly, blocking up to 90% of the light. Other switchable windows are designed to provide privacy by changing from transparent to translucent, but are not effective at saving energy.[30,31] Another proposed window design includes sensors that automatically raise or lower a

Fig. 9 Energy Star's four regions used for window rating program.

Table 1 Energy Star window criteria

Climate zone	*U*-factor	Solar heat gain coefficient (SHGC)	
Northern	≤0.35	Any	
North/Central	≤0.40	≤0.55	
South/Central	≤0.40	≤0.40	Prescriptive
	≤0.41	≤0.36	Equivalent performance (excluding CA)
	≤0.42	≤0.31	
	≤0.43	≤0.24	
Southern	≤0.65	≤0.40	Prescriptive
	≤0.66	≤0.39	Equivalent performance
	≤0.67		
	≤0.68	≤0.38	
	≤0.69	≤0.037	
	≤0.70		
	≤0.71	≤0.36	
	≤0.72	≤0.35	
	≤0.73		
	≤0.74	≤0.34	
	≤0.75	≤0.33	

blind enclosed between two panes of glass based upon the outdoor temperature and solar radiation. This design admits solar radiation when it will help heat the house, but lowers the blind to block solar radiation when it will increase the air conditioning load.[32]

CONCLUSION

Walls and windows are often selected to achieve a desired appearance. Considering today's emphasis on energy conservation and overall sustainability, it is important to consider their thermal characteristics as well. Selecting more energy-efficient walls and windows will permit the building designer to specify a smaller heating and cooling system, so that often the total building cost is little more than that of a standard building. When you consider the reduced cost of heating and cooling the building during its lifetime, any added investment during construction is returned many times over.

OTHER USEFUL GUIDES

Graphic Guide to Frame Construction: Walls, by *Rob Thallon*, Published by Taunton Press, 2000, ISBN 1-56158-353-7, # 070470 (http://www.taunton.com/finehomebuilding/pdf/Framing%20Walls.pdf)
http://www.taunton.com/finehomebuilding/pdf/Grading%20and%20Drainage.pdf

REFERENCES

1. http://www.energycodes.gov/ (accessed February 12, 2007).
2. Ontario Association of Architects, *The Rain Screen Wall*, found at http://www.cmhc-schl.gc.ca/en/inpr/bude/himu/coedar/loader.ffm?url=/commonspot/security/getfile.cfm&PageID=70139, other related publications from the Canadian Mortgage Housing C… available at http://www.cmhc-schl.gc.ca/en/ (accessed February 12, 2007).
3. Toolbase Services, a consortium including the U.S. Department of Housing and Urban Development and the National Association of Home Builders (NAHB) provides a wealth of information about current and future building materials and housing systems, NAHB Research Center, Maryland, 2005, found at www.toolbase.org (accessed February 13, 2007).
4. Kosny, J.; Kossecka, E.; Desjarlais, A.O.; Christian, J.E. In *Dynamic Thermal Performance of Concrete and Masonry Walls*, Thermal Envelopes VII, Conference Proceedings, Clearwater Beach, FL, December 1998; ASHRAE Special Publications, 1998.
5. Lippke, B.; Wilson, J.; Perez-Garcia, J.; Bowyer, J.; Meil, J. CORRIM: life-cycle environmental performance of renewable building materials. Forest Prod. J. **2004**, *54* (6), 8–19 http://www.corrim.org/reports/pdfs/FPJ_Sept2004.pdf, (accessed February 13, 2007).
6. Brown, M.A.; Southworth, F.; Stovall, T.K. *Towards a Climate-Friendly Buildt Environment*; Pew Center on Global Climate Change: Arlington, VA, 2005; available at

Util–Wall

http://www.pewclimate.org/document.cfm?documentID=469 (accessed February 13, 2007).
7. Stovall, T. K., and Karagiozis, A., *Airflow in the Ventilation Space Behind a Rain Screen Wall*, Thermal Envelopes IX, Conference Proceedings, Clearwater Beach, FL, December 2004; ASHRAE Special Publications, 2004.
8. U.S. Department of Energy, Technology Fact Sheet: Wall Insulation, October, 2000, http://www.eere.energy.gov/buildings/info/documents/pdfs/26451.pdf (accessed February 13, 2007).
9. Kosny, J.; Mohiuddin, S. A. *Interactive Internet-Based Building Envelope Materials Database for Whole-building Energy Simulation Programs*, Thermal Envelopes IX, Conference Proceedings, Clearwater Beach, FL, December 2004; ASHRAE Special Publications, 2004.
10. Toolbase Services, a consortium including the U.S. Department of Housing and Urban Development and the National Association of Home Builders (NAHB) provides a wealth of information about current and future building materials and housing systems, NAHB Research Center, Maryland, 2005, found at www.toolbase.org
11. Kosny, J. *Steel Stud Walls; Breaking a Thermal Bridge*. Home Energy July 2001, http://www.homeenergy.org/article_preview.php?id=249&article_title=Steel_Stud_Walls:_-Breaking_the_Thermal_Bridge (accessed February 13, 2007).
12. ASHRAE Handbook of Fundamentals, 2005 Chapter 25.
13. Kosny, J.; Petrie, T.; Gawin, D.; Childs, P.; Desjarlais, A.; Christian, J. Thermal Mass—Energy Savings Potential in Residential Buildings, August 2001, http://www.ornl.gov/sci/roofs+walls/research/detailed_papers/thermal/index.html (accessed February 13, 2007).
14. Advanced Wall Systems Hotbox Test *R*-value Database, Oak Ridge National Laboratory, http://www.ornl.gov/sci/roofs+walls/AWT/Ref/TechHome.htm
15. NAHB Research Center, ToolBase Services, "PATH Technology Inventory, Precast Concrete Foundation and Wall Panels," Upper Marlboro, MD, 2005, http://www.toolbase.org/Technology-Inventory/Foundations/precast-concrete-panels (accessed February 13, 2007).
16. Precast/Prestressed Concrete Institute, "Single Family Housing," http://www.pci.org/markets/projects/single.cfm (accessed February 13, 2007).
17. Thermal conductivity estimated considering the reported thermal conductivity of unfired clay bricks found at http://www.constructionresources.com/products/ and the measured thermal conductivity of a wall built from rammed earth bricks reported at http://www.ornl.gov/sci/roofs+walls/AWT/Ref/TechHome.htm (accessed February 13, 2007).
18. Christian, J.E.; Desjarlais, A.O.; Stovall, T.K. In *Straw Bale Wall Hot Box Test Results and Analysis*, Conference Proceedings, Clearwater Beach, FL, December, 1999; Thermal Performance of the Exterior Envelopes of Buildings VII, 1999; 275–285.
19. Carmody, J.; Selkowitz, S.; Heschong, L. *Residential Windows, A Guide to New Technologies and Energy Performance*; W.W. Norton & Company: New York, 1996.
20. Joint Center for Housing Studies of Harvard University, *Remodeling, Measuring the Benefits of Home Remodeling*; Harvard University: Cambridge, MA, 2003.
21. U.S. Department of Energy, 2004 Buildings Energy Databook, http://buildingsdatabook.eren.doe.gov/default.asp
22. U.S. Department of Energy, A Consumer's Guide to Energy Efficiency and Renewable Energy, Storm Windows, http://www.eere.energy.gov/consumer/your_home/windows_doors_skylights/index.cfm/mytopic=13490
23. Klems, J.H. In *Measured Winter Performance of Storm Windows*, ASHRAE 2003 Meeting, Kansas City, MO, June 28–July 2, 2003; LBNL-51453.
24. Turrell, C., Storm windows Save Energy, Home Energy Magazine Online, July/August 2000, http://homeenergy.org/archive/hem.dis.anl.gov/eehem/00/000711.html
25. The NFRC has developed a number of informative fact sheets regarding window issues. You can find them at www.nfrc.org/factsheets.aspx (accessed February 13, 2007).
26. www.energystar.gov/index.cfm?c=windows_doors.pr_windows
27. The *U*-factors used by the NFRC and the Energy Star program have units of Btu/h-ft^2-°F.
28. www.nfrc.org/label.aspx (accessed February 13, 2007).
29. http://www.energystar.gov/index.cfm?c=windows_doors.win_elig
30. Fanjoy, R. Smart Windows Ready to Graduate, HGTV Pro.com, found at http://www.hgtvpro.com/hpro/dj_construction/article/0,2619,HPRO_20156_3679648,00.html (accessed February 13, 2007).
31. Toolbase Services, a consortium including the U.S. Department of Housing and Urban Development and the National Association of Home Builders (NAHB) provides a wealth of information about current and future building materials and housing systems, NAHB Research Center, Maryland, 2005, found at www.toolbase.org
32. Kohler, C.; Goudey, H.; Arasteh, D. A First-Generation Prototype Dynamic Residential Window, LBNL-56075, Lawrence Berkeley National Laboratory, Berkeley, CL, October 2004, available at http://www-library.lbl.gov/docs/LBNL/560/75/PDF/LBNL-56075.pdf (accessed February 13, 2007).

Waste Fuels

Robert J. Tidona
RMT, Incorporated, Plymouth Meeting, Pennsylvania, U.S.A.

Abstract
Opportunity fuels are those wastes or process byproducts that have significant energy content and could be used to provide energy to generate electricity but have not traditionally been used for that purpose. Burgeoning interest in the use of opportunity fuels to offset purchased traditional fossil fuels has focused on the combustion, material handling, and environmental permitting challenges associated with their exploitation. A number of industries have been taking advantage of opportunity fuels since long before that moniker was coined. Other industries are relative newcomers to the field. Some industry analysts have indicated the potential for as much as 100 GW of electric generation from opportunity fuels associated with distributed generation facilities. Although such estimates are not unfounded, a number of challenges are associated with a potential application of the use of opportunity fuels. These challenges include defining the fuel sufficiently for equipment vendors and regulatory agencies. Characteristics to be established are fuel particle size, shape, and propensity to bridge, agglomerate, stick to equipment surfaces, leach, or contain hazardous components. Other important considerations are the ease and cost of transporting the fuel (if required); its ability to be co-fired with other traditional or waste fuels; the ability to obtain performance guarantees from boiler/furnace/steam generator vendors with the desired opportunity material; and pollutant emissions from combustion, gasification, and handling. The answers to these issues, along with the typical design issues revolving around complex multifuel steam plants with on-site electric generation, must be obtained in the course of any feasibility analysis of opportunity fuels.

INTRODUCTION

The price of natural gas has been steadily rising throughout the 1990s and the first half-decade of the new millennium. From a benchmark NYNEX price averaging about $2–$4/MMBtu in the early 1990s to a 2006 level of approximately $10–$14/MMBtu, the spot market price has doubled over the past decade. Perhaps even more significant, the *perception* that we are in the midst of a significant upward trend in natural gas prices has grown increasingly strong. Increased demand from electric utilities, which have become more dependent upon this relatively clean fossil fuel over the past decade to fuel their gas turbine peaking plants and natural gas-fired steam boilers, is creating "shortages" at peak usage times, thereby putting upward pressure on the price of gas. The upward spike in the price of natural gas is illustrated in Fig. 1.

The relative ease of environmental permitting of natural gas or light distillate oil combustion devices, whether for electric generation or not, has made these fuels the popular choice for industrial, commercial, and utility boilers and furnaces nationwide. In many cases, to obtain a permit in a timely fashion, alternative fuels that might be more economically attractive, but are as yet unproven for specific applications, often lose out to these more traditional clean fossil fuels. Natural gas and distillate oil have become the "fuels of least resistance" from an environmental permitting standpoint, making them more valuable and driving up prices.

The upward trend in fossil-fuel prices, exacerbated by the shortages created by the Gulf hurricanes of 2005, has driven many companies to begin seriously considering alternative and opportunity fuels to supplement or replace their current fossil fuels. This article examines the current status of these fuels and their potential.

Because this topic crosses many disciplinary lines, a number of terms are defined below for purposes of the ensuing discussion.

EXAMPLES OF ALTERNATE AND OPPORTUNITY FUELS

Alternate fuel is a general term that can apply to any fuel not normally used by a process, boiler, or furnace but that can be used to provide energy. These fuels can be utilized either with or without changes to burner, fuel, and fuel residue handing, or pollution control systems to supplement or replace the use of fossil fuels. Numerous

Keywords: Alternate fuel; Opportunity fuel; Waste fuel; Industrial energy recovery; Municipal waste; Solid waste; Cogeneration; Distributed generation.

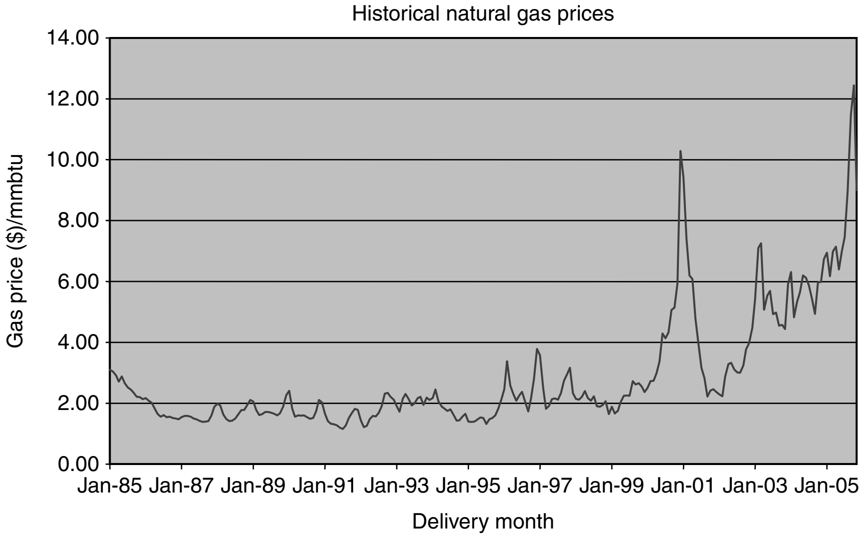


Fig. 1 Historical natural gas prices.
Source: From RMT, Incorporated internal report, December 2005.

examples of alternative and opportunity fuels are listed below:

- Agricultural wastes: corn cobs, cotton seed hulls, rice hulls, peanut shells, sugar cane waste, coconut husks.
- Anaerobic digester gas: gas consisting of approximately two-thirds methane and one-third carbon dioxide, along with a few percent nitrogen and small quantities of other gases, including oxygen, hydrogen, and hydrogen sulfide. This gas is produced by the anaerobic digestion process, which is often associated with sewage treatment, animal waste, vegetable oil, or alcohol mills. Anaerobic digestion is the process of using microorganisms that live only in the absence of oxygen to decompose organic materials into more valuable, or more easily disposed, components. The digestion process can also be used for synthetic natural gas from organic wastes or from algae.
- Bagasse: the fibrous material remaining after the extraction of the juice from sugar cane. It is used as a fuel and as a mix in making lightweight refractories.
- Black liquor: from pulp and paper plants, the organic residue from the wood digestion process that is burned in recovery boilers to eliminate the wood organics and to recover sodium sulfate for reuse in the wood digestion process. It has a heating value of approximately 4000 Btu/lb.
- Blast furnace gas: from steel mills, a gas, low in heat content (typically 100 Btu/ft^3 or less), recovered from a blast furnace as a byproduct and used as a fuel.
- Carpet waste: surplus cuttings and rejected product from manufacturing, as well as postconsumer carpet waste.
- Coalbed methane: methane that diffuses out of seams in coal mines and is channeled to the surface by mine shafts.
- Coke oven gas: obtained as a byproduct of coke production processes (mostly hydrogen and methane) with a heating value typically just under 600 Btu/ft^3. Coke is the solid, combustible residue left after the destructive distillation of coal or crude petroleum.
- Food processing wastes: include animal carcasses, animal fats, citrus rinds, peanut shells, and corn husks.
- Industrial VOCs: volatile organic chemicals, usually emitted in the manufacture of products.
- Industrial organic sludges: heavy liquids with integral suspended and possible dissolved solid phases.
- Landfill gas: approximately 50% methane and 45% CO_2, generated by the anaerobic decay of organic materials.
- Municipal solid waste (MSW): includes trash; yard wastes; garbage; household waste (including paper, wood, and carpet waste); textile waste; plastic and other synthetic packaging materials; and minor amounts of metals, soil, and other noncombustibles.
- Petroleum coke: coal (carbon black) that has been heated and partially oxidized.
- Refuse-derived fuel (RDF): fuel made from municipal solid waste, typically in pellet form.
- Tape manufacturing waste: depending on the product line, can be paper- or cloth-backed waste, either coated with adhesive or noncoated, and can include cardboard and plastic packaging materials, unusable end rolls, trimmings, waste from web breaks, and off-specification product.

- Textile wastes: cuttings and rejected product from manufacturing.
- Tires: whole, chipped, shredded postconsumer waste.
- Tire-derived fuel: fuel made from postconsumer tires, usually in pellet form.
- Used oil: defined in Title 40 of the Code of Federal Regulations (CFR) Part 279, Standards for the Management of Used Oil, as "any oil that has been refined from crude oil, or any synthetic oil, that has been used and as a result of such use is contaminated by physical or chemical impurities." Used lubricants, hydraulic fluids, and heat transfer fluids are examples. Used oil typically is contaminated or mixed with dirt, fine particles, water, or chemicals, all of which affect the performance of the oil and eventually render it unusable. By itself, it is not a hazardous waste, even though it may exhibit hazardous waste characteristics.
- Wastewater treatment sludge: heavy liquid with integral suspended and possible dissolved solid phases, usually dried by mechanical or heat exchanger method.
- Wood products industry:
  - Sanding/finishing dust: waste material from sanding or other finishing process, such as grinding and sawing
  - Hogged (wood) fuel: wood logs that have been reduced in size in a shredding machine called a hogger
  - Bark: stripped from the surface of logs, usually with a high moisture content (often at or above 50%).

A partial summary of the key characteristics of opportunity fuels is contained in Table 1.

MATERIAL HANDLING AND WASTE DISPOSAL

Blending of waste materials with other wastes or with conventional fuels can be the key to ultimate success in utilizing opportunity fuels. Blending can take place either en route to day storage or just prior to introduction into the energy conversion device. The target attributes of the blend include consistency of thermal energy content, particle sizing, material handling characteristics, bulk density, stickiness, and ash content.

Typical sizing of fuel specified by stoker boiler vendors is 2 in. maximum particle size in any dimension, with uniform sizing generally desired. An occasional large particle can be tolerated, provided that the material is relatively unlikely to agglomerate or accumulate in pipes or ducts. A very light material can be expensive to store due to its low bulk density and correspondingly large required storage vessels. A very heavy material can cause problems with elevating or pneumatic conveyors.

Standard disposal of solid waste materials is by landfill or municipal incinerator. The costs of these disposal options may range from $10 to $40 per ton of waste for landfilling and up to $80 per ton for incineration. Often, these costs can be eliminated in large part by co-firing the wastes in a boiler or furnace. Only the remaining noncombustible ash components and any residue remaining from the neutralization of acid gases in the boiler exhaust gases (see the next section for more details) will still need to be landfilled.

ENVIRONMENTAL CONSIDERATIONS

Combustion Processes

Federal, state, and in some cases county and/or city regulations that apply in most cases include Title 40 of the CFR Part 60 New Source Performance Standards (NSPS); 40 CFR Part 63 National Emission Standards for Hazardous Air Pollutants (NESHAP), specifically the Industrial, Commercial, and Institutional Boiler and Process Heater Maximum Achievable Control Technology (ICIB/PH MACT) rule; and process-based MACT standards.

- Gaseous emissions: NO_x, SO_2, CO, HCl, VOC, Hg.

Note that SO_2, HCl, and Hg emissions are primarily associated with initial concentrations of sulfur, chlorine, and mercury in the waste material prior to thermal processing, whereas NO_x, CO, and VOC occur due to the presence (or lack thereof) of oxygen and the mixing, or lack thereof, of fuel and air in the combustion process.

- Particulate emissions:
  - Total particulate matter (PM)
  - Total selected metals (TSM): includes 6 metals that are known health threats.
- Opacity.

NonCombustion Processes

Gasification

Gasification is gaining interest as an alternative to combustion-based energy generation, particularly among industrial concerns. The advantage of gasification is that a relatively clean, gaseous fuel may be generated from a solid waste stream, eliminating the need for expensive particulate control equipment at the downstream end of the process. In addition, for the production of electricity, gasification may permit direct conversion to power, averting the need for a steam generator and turbine. Some cleanup of the gas is typically required prior to being introduced into a burner. Permitting considerations are dependent upon the particular application and location but

Table 1 Opportunity fuel performance matrix

Opportunity fuel	Availability	Heating value	Fuel cost	Equipment cost	Emissions/ environment	Combined heat and power (CHP) potential	Rating	Limitations
Anaerobic digester gas	●	◑	●	◑	●	●	5.0	Need anaerobic digester
Biomass gas	●	◑	◑	○	●	●	4.0	Gasifiers extremely expensive
Black liquor	○	◑	●	◑	◑	◑	3.0	Most BL already used up by mills
Blast furnace gas	○	○	●	◑	◑	○	2.0	Limited availability, low Btu
Coalbed methane	◑	●	●	●	●	◑	5.0	Coal mines—lack CHP demand
Coke oven gas	○	◑	●	◑	◑	◑	3.0	Availability—most already used
Crop residues	◑	◑	○	◑	●	◑	3.0	Difficulty in gathering/transport
Food processing waste	◑	◑	●	◑	●	◑	4.0	Limited market, broad category
Ethanol	●	◑	◑	◑	●	◑	4.0	Currently only used for vehicles
Industrial volatile organic chemicals (VOC's)	○	○	◑	◑	◑	◑	2.0	Must be used w/ NG turbine
Landfill gas	●	◑	●	◑	●	◑	4.5	Landfills—little demand for CHP
MSW/Refuse-derived fuels (RDF)	●	○	●	○	◑	◑	3.0	Low heating value, contaminants
Orimulsion	○	●	◑	◑	◑	◑	2.5	Orimulsion not available in U.S.
Petroleum coke	●	●	●	◑	○	○	3.5	Many contaminants; large apps
Sludge waste	●	○	●	○	◑	○	2.5	Low heating value, contaminants
Textile waste	◑	◑	●	◑	◑	○	3.0	Must be cofired; larger apps
Tire-derived fuel	●	●	◑	◑	◑	◑	4.0	Best suited for large apps
Wellhead gas	◑	●	●	◑	●	◑	4.5	Oil/gas wells—no CHP demand
Wood (forest residues)	●	◑	◑	◑	●	◑	4.0	Fuel can be expensive
Wood waste	●	◑	●	◑	◑	◑	4.5	Waste may have contaminants

Key: ●, excellent/not an issue; ◑, average/could become an issue; ○, poor/major issue.
Source: From Resource Dynamics Corporation (see Refs. 1 and 2).

can be somewhat reduced from combustion-based equipment. A more detailed discussion of gasification technologies may be found in "Gasifiers."

Gasification, however, is not a universally applicable technology and can be expensive, relative to combustion. Further, there is limited full-scale operating experience with this technology.

Anaerobic Digestion

Anaerobic digestion is the process of using microorganisms that live only in the absence of oxygen to decompose organic materials into more valuable or more easily disposed-of components. The digestion process can also be used for synthetic natural gas from organic wastes or from algae. As with gasification processes, because anaerobic digestion does not involve combustion, the time and effort spent meeting environmental permitting requirements tends to be reduced.

TECHNOLOGIES

Retrofit Applications

Existing Solid Fuel Boilers

Many of the best opportunities for energy recovery are to be found in industry. The pulp-and-paper and petroleum refining industries have long utilized numerous waste streams for energy. There are more than 7000 MW of biomass-fired energy systems in place in the United States. Portland cement plants have been co-firing tires in their kilns since the 1970s.

Co-firing is a promising technology for burning opportunity fuels within conventional coal-fired boilers. Fig. 2 shows an example of a commercially available burner for this application.

Waste Heat Recovery Steam Generators

Often, wastes are burned in incinerators, and the hot flue gas is exhausted to the atmosphere. The use of waste heat boilers and steam generators is gaining popularity as traditional fuel costs rise and state-of-the-art exchanger design and control systems provide added margins of safety.

Burner/Engine Redesign

Occasionally, an excellent opportunity for the reclamation of the energy potential from wastes may be obtained by a redesign of an existing burner or engine. These opportunities include the use of bagasse in Combined Heat and Power (CHP) plants and other relatively high-Btu waste gas streams, such as landfill gas and coke oven gas.

Sludge Drying

Some chemical or biological sludges contain large quantities of moisture after initial dewatering steps. It is sometimes cost effective to reduce the moisture levels in these sludges prior to burning them in a boiler or furnace. Many dewatering processes are available. They include physical or mechanical means such as centrifuges or filter presses, as well as thermal drying systems. Typically, the mechanical-type dewatering systems are the initial step in sludge drying. The high heat transfer rates possible in specialized thermal drying equipment (such as hollow screw or paddle type dryers) can provide enhanced drying capability, resulting in relatively high solids content, thus further reducing the need for fuel to evaporate the moisture in a combustion process. During the process of drying some materials, proper consideration must be given to the potential for the development of a "sticky" phase in which the material becomes non-Newtonian and can agglomerate or plug the handling system.

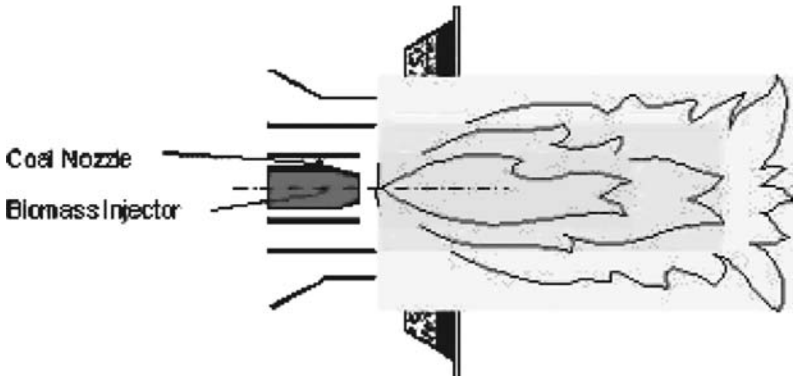


Fig. 2 Biomass co-firing through air. Source: From Foster Wheeler Corporation (see Ref. 3).

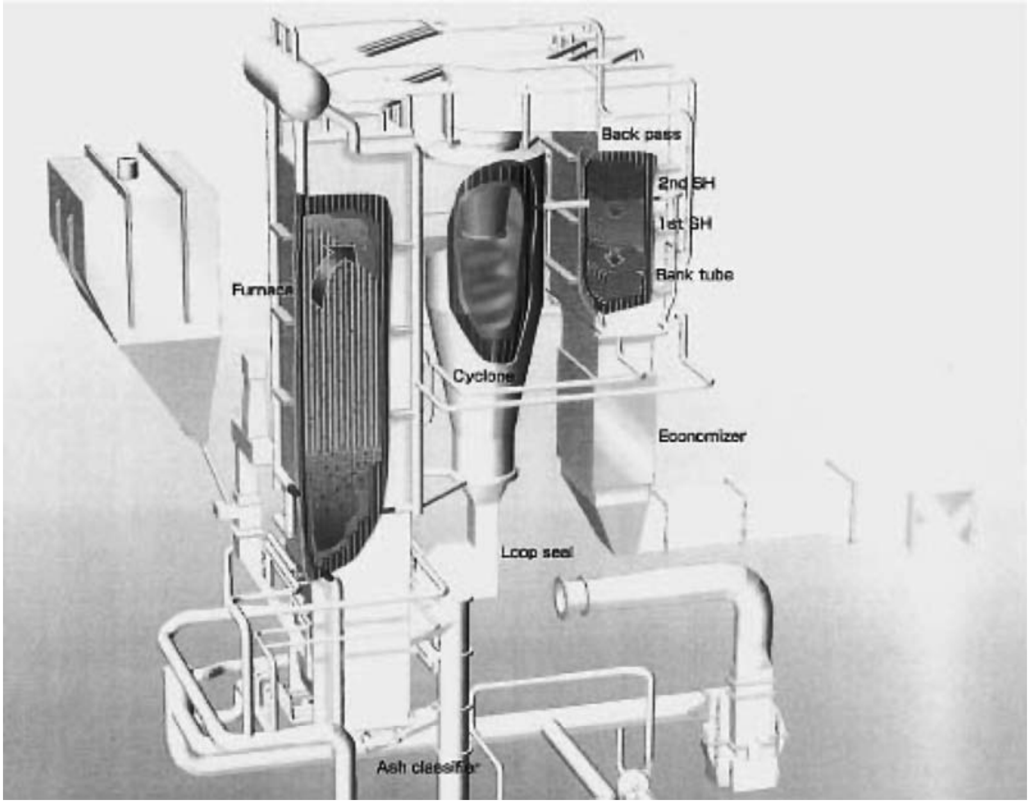


Fig. 3 Coal/biomass—fired circulating fluid bed boiler.
Source: From Takuma Company, Ltd. (see Ref. 4).

New Equipment

Circulating Fluid Bed (CFB) and Bubbling Fluid Bed (BFB) Boilers

Fluidized bed boilers have been in use for many years. They are well known for their capability to burn diverse materials efficiently and with low emissions. Compared with the older stoker or underfed boiler designs of the past, they can achieve greater burnout of the carbonaceous feed materials at relatively low emission rates of NO_x and PM.

As their name implies, Circulating Fluid Bed (CFB) boilers achieve lower and more uniform combustion-zone temperatures and higher heat transfer rates by recirculating the bed material. Consequently, CFB boilers generally have lower NO_x emissions and can burn many types of fuels. Their chief advantage over the lower-velocity, fixed fluid bed boilers is that they can achieve higher energy densities and, consequently, smaller footprints. It is expected that these types of boilers will have an efficiency and flexibility advantage for larger steam and electric generation applications.

Some commercial suppliers of CFB boilers are

- Foster Wheeler
- Babcock and Wilcox
- Kvaerner Power
- Alstom
- Takuma

Fig. 3 shows an example of a coal/biomass circulating fluidized bed boiler.

Bubbling fluid bed boilers (BFBs) have been in use longer than CFBs and are more suited for lower-heating-value materials, such as waste sludges and biomass fuels containing large amounts of moisture. Because they do not recirculate the bed material as in CFBs, they tend to be somewhat lower in cost for the same duty requirement, though their physical size is larger. Turndown is expected to be greater for BFBs than CFBs.

Examples of suppliers of BFB boilers are

- Babcock and Wilcox
- Kvaerner Power
- Austrian Energy
- Energy Products of Idaho

Gasifiers

Due to rising natural gas and distillate oil prices, there is intense interest in the United States regarding alternate fuels. Where space or air pollution control constraints make the application of large combustion devices such as CFBs and all the ancillary equipment accompanying them problematic, gasifiers offer energy recovery that can be permitted in a relatively short period and that, depending on the syngas cleanup requirements, requires a relatively small installed footprint.

The following Web site provides an excellent description of the major gasification processes: www.eere.energy.gov/biomass/large_scale_gasification.html (Ref. 5).

There are several widely used process designs for biomass gasification. In staged steam reformation with a fluidized-bed reactor, the biomass is first pyrolyzed in the absence of oxygen; then the pyrolysis vapors subsequently reform to synthesis gas, with steam providing added hydrogen, as well as the proper amount of oxygen and process heat that comes from burning the char. With a screw auger reactor, moisture (and oxygen) are introduced at the pyrolysis stage, and process heat comes from burning some of the gas produced in the latter stage. In entrained flow reformation, both external steam and air are introduced in a single-stage gasification reactor. Partial oxidation gasification uses pure oxygen, with no steam, to provide the proper amount of oxygen. (Using air instead of oxygen, as in small modular uses, yields producer gas including nitrogen oxides rather than synthesis gas.)

Fig. 4 shows a simplified process diagram of a fluidized bed gasifier.

Commercial suppliers of gasifiers include

- Energy Products of Idaho (EPI)
- Pyromex
- Primenergy
- Chiptec
- Interstate Waste Technologies
- Omnifuels Technology, Inc.
- Precision Energy Services
- PRM Energy Systems, Inc.
- Thermogenics, Inc.

Anaerobic Digesters

Anaerobic digestion is gaining in popularity in the United States due to its reduced processing cycle times and energy requirements. This process eliminates the need for expensive pumps required in aerobic (oxygenated) processes.

Anaerobic sludge digestion is the destruction of biological solids using bacteria that function in the absence of oxygen. This process produces methane gas, which can be used as an energy source and can make anaerobic digestion more economically attractive than aerobic digestion. In essence, the larger the treatment plant, the greater the economic incentive to use anaerobic digestion. Anaerobic digestion is considerably more difficult to operate than aerobic digestion, however. As such, the decision to use anaerobic digestion must take into consideration the operational capability of the installation.

There are two conventional processes: mesophilic, which takes place at ambient temperatures typically between 20 and 40°C, and thermophilic, which takes place at elevated temperatures, typically up to 70°C. The residence time in a digester varies with the type and amount of feed material and the temperature. In the case of mesophilic digestion, residence time may be between 15 and 30 days; the thermophillic process is usually faster, requiring only about 2 weeks to complete, but thermophilic digestion is more expensive, requires more energy, and is less stable than the mesophillic process.

Typically, two types of organisms are found in digester sludge:

- Saprophytic bacteria, which are acid formers that break down complex solids to volatile acids
- Methane fermenters, which produce methane, carbon dioxide, and inert materials from the volatile acids that are produced by the saprophytic bacteria

The efficiency of the digesters is typically measured by the amount of volatile solids reduction between the raw sludge and the digested sludge. Many variables affect the volatile solids reduction in the digestion process. The major variables are sludge type, digestion time, digestion temperature, and mixing.

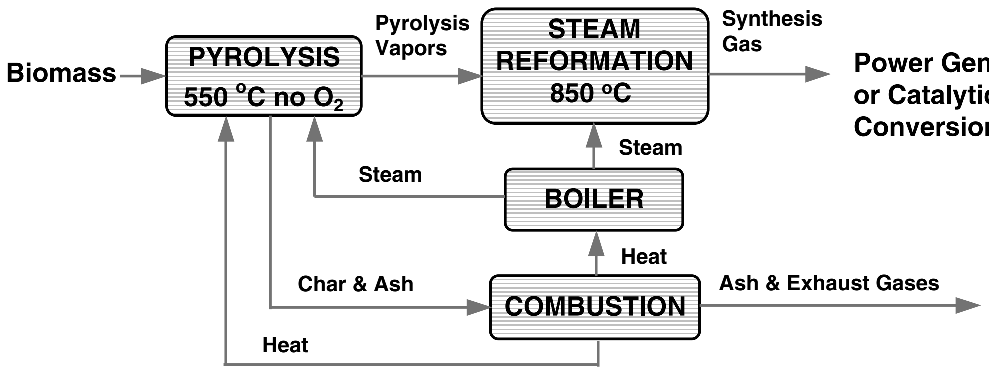


Fig. 4 Biomass gasification via staged steam reformation with a fluidized bed gasifier.

Was–Wat

Applications for digesters fall into two major categories: wastewater treatment and farm-animal wastes.

Funding for anaerobic digester projects is available in some states. Details of one such program can be found on the New York State Energy Research and Development Administration (NYSERDA) Web site: www.nyserda.org/programs/pdfs/digestergrantlist.pdf (Ref. 6).

Commercial suppliers of anaerobic digesters include

- Arrow Ecology
- CCI US Corporation
- EcoCorp, Inc.
- Onsite Power Systems

POTENTIAL AND ECONOMICS

Reduced-Cost Alternate and Opportunity Fuels

Commercial Waste Fuel Suppliers

Biomass fuels are available in all parts of the country, particularly where the forest-products industry has a presence. Commercial suppliers are finding many new customers that are seeking alternatives to higher-priced traditional fossil fuels for boiler plants and power generation facilities. Mixed wood fuels from pulping, hogging, and debarking operations, as well as sanderdust, sawdust, and furniture manufacturing waste, can be sourced through a number of commercial suppliers.

The relative costs of many opportunity fuels may be estimated through information readily available on the World Wide Web at sites like the North Dakota State University's Fuel Cost Comparison site: www.ext.nodak.edu/extpubs/ageng/structu/ae1015.htm (Ref. 7).

Recycling Centers

Recycling of many manufacturing and postconsumer wastes for use as opportunity fuels has begun in earnest. As an example, nearly 5 billion lb/yr of carpet waste is generated by the carpet manufacturing industry. Much of this waste material is produced in Georgia. Commercial "reprocessors" of this waste material are pelletizing, baling, compressing, and gasifying it for its ultimate energy utilization.

Disposal Cost Reduction

Landfill Cost

Landfill of wastes can cost in the range of $10–$40 per ton. Typically, the greater the hauling distance, the greater the cost of disposal. This operating expense, coupled with the loss of usable thermal energy ("opportunity cost"), has often been ignored in an era of relatively inexpensive fossil fuels. This is no longer the case in many areas. Landfills themselves are aggressively seeking cost mitigation through energy utilization of landfill gas methane. As landfill costs climb due to the diminishing available land, increased environmental restrictions on groundwater cleanliness, and increased transportation costs as landfills are sited more remotely from population centers, conversion of waste to energy (WTE) is gaining more attention.

Trash Collection and Hauling

Just as landfill costs are spurring interest in WTE, increased trash hauling and collection costs are adding fuel to the fire. Industries are considering their options when it comes to paying rental costs for numerous waste containers on their manufacturing sites and the emptying of those containers. Such costs can run thousands of dollars per day at many manufacturing facilities.

Combined Heat and Power (CHP) for Increased Efficiency

Thermal Load Following (Bottoming) Cycle

An example of a thermal load following WTE is a process steam boiler with a backpressure turbine generator. This type of facility typically uses steam throughput rates from about 40,000 to several hundred thousand pounds of steam per hour. Steam pressures typically range from 150 psig to several thousand pounds per square inch gauge. Electric output can range from 100 kW up to about 20 MW.

Limitations include the necessity to use all steam or vent the excess that cannot be used by the process or the turbine. Facilities with greatly varying steam loads can experience problems with upset conditions caused when loads change rapidly and without warning. The turbines can trip offline, and process steam pressures and temperatures may fluctuate.

These problems can be alleviated by the use of a condensing turbine at the cost of additional capital. Larger systems with electric generation capacities of several MW or more will usually have a condensing turbine. The flexibility offered by the condensing turbine allows a boiler to run base loaded, which permits the use of large fluidized bed boilers designed for a multitude of fuels. Extraction turbines allow the tapping of steam slipstreams at various pressures for process use.

Electric Load Following (Topping) Cycle

An example of an electric load following cycle is a central steam boiler plant, which generates electricity to supply its users' base electrical loads via a condensing turbine.

The steam that is used in the turbine can be reused for process or heating needs at the exhaust pressure of the turbine or at intermediate pressures extracted from the turbine at certain points in the expansion from turbine inlet to turbine exhaust pressure.

Advantages include the ability to accommodate variable steam loads simply by condensing more or less steam. This also provides flexibility during startup and transient steam load conditions, and can reduce fluctuations in steam flow, temperature, and pressure to process users at such times.

The disadvantages are added initial cost and complication of the steam flow system and turbine due to the addition of the condenser; increased turbine manufacturing cost; and increased piping, wiring, and installation costs.

Sizes of condensing turbines range from about 5 to 100 MW of electrical output.

The potential for generation from CHP plants burning opportunity fuels has been estimated at more than 100 GW.

Table 2 shows a chart comparing estimated capital and operating and maintenance costs to generate electricity utilizing several opportunity fuels.

Wood fuels in particular have been studied for their potential use in CHP, including both thermal and electric load following applications. Table 3 illustrates the costs associated with their construction, operation, and maintenance.

Integrated Coal or Waste Gasification/Combined Cycle (IGCC) Power Plants

An alternative to conventional boilers in which fuels are burned to release energy that is then used to make steam is gasification wherein a solid fuel is converted to a gaseous fuel stream by means of thermal energy addition—generally in an oxygen-deficient atmosphere.

The integrated coal gasification/combined cycle is one version of this concept. There are many others, and the concept is growing in popularity among operators of large industrial and institutional boiler plants because of the relative ease of permitting this type of facility compared with installing a traditional boiler.

The advantage of this type of system is generally regarded as being its ability to convert coal or a waste byproduct (such as a solid waste or sludge) to usable energy, typically in the form of steam and/or electricity.

Project Funding Opportunities

Biomass Programs

- U.S. Department of Energy (DOE) grant program:
  The DOE provides grants for projects that have the objective of furthering the use of biomass fuels.
- U.S. Department of Agriculture loan guarantee program:
  The USDA has a loan guarantee program for projects that expand the use of agricultural byproducts as energy sources.
- State-funded programs:
  The following is a useful link to a U.S. DOE Energy Efficiency and Renewable Energy (EERE) Web site that gives a great deal of information about programs run by individual states: www.eere.energy.gov/state_energy_program (Ref. 8).

Renewable Energy Resources

The general category of renewable energy resources includes biomass and other opportunity fuels, in addition to traditional solar and wind energy sources.

Fig. 5 shows two possible projections for the future composition of the U.S. renewable energy portfolio.

Fig. 6 is a projection of the cumulative savings of natural gas resulting from the increasing use of renewable energy resources.

Cogeneration

The DOE supports a series of regional CHP Application Centers to provide application assistance, technology information, and education to architects and engineering companies, energy services companies, and building owners in eight states (Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, and Wisconsin).

DOE's Office of Industrial Technology (OIT) coordinates two programs—the Combustion Program and Steam Challenge Program—to support the use of CHP technologies. The Combustion Program's goal is to foster research and development opportunities for combustion systems used to generate steam and heat for manufacturing processes, and the heat-processing materials from metals to chemical feedstocks. The Steam Challenge Program provides technical assistance to industrial customers to improve the efficiency of their steam plants and increase awareness of benefits from cogeneration systems. OIT predicts opportunities for future CHP development for any industrial business where boilers are installed.

The Office of Distributed Energy Resources coordinates the CHP Initiative to raise awareness of the energy, economic, and environmental benefits of CHP and to highlight barriers that limit its increased implementation.

Greenhouse Gas Emission Reduction

Projects in developing countries that seek to reduce the emissions of greenhouse gases (GHGs), including CO_2, methane, nitrous oxide, hydrofluorocarbons (HFCs), and SF_6 can obtain financial support through a Kyoto Protocol

Table 2 Equipment and maintenance average costs [not including combined heat and power (CHP) equipment]

Fuel	Type of cost	Steam turbine[a]($)	Gas turbine($)	Recip engine($)	Microturbine($)	Fuel cell($)	Stirling engine($)
Anaerobic digester gas[b]	Equipment ($/kW)	2150–3500	1800–3600	1900–3200	2650–4400	4800–7500	2400–3500
	Maintenance ($/kWh)	0.007–0.022	0.008–0.021	0.015–0.043	0.025–0.035	0.21–0.043	0.009–0.013
Biomass gas[c]	Equipment ($/kW)	1700–2800	1300–2550	1500–2550	2150–3500	4100–6500	2100–3000
	Maintenance ($/kWh)	0.007–0.022	0.005–0.016	0.01–0.029	0.017–0.027	0.017–0.038	0.009–0.015
Coalbed methane	Equipment ($/kW)	1000–1600	600–1400	800–1400	1400–2300	3500–5500	1500–2000
	Maintenance ($/kWh)	0.005–0.015	0.004–0.01	0.008–0.022	0.015–0.02	0.015–0.03	0.008–0.01
Landfill gas	Equipment ($/kW)	1250–2000	900–2100	1000–1700	1750–2900	3900–6000	1500–2000
	Maintenance ($/kWh)	0.016–0.019	0.007–0.018	0.014–0.04	0.024–0.032	0.02–0.04	0.008–0.01
Tire-derived fuel	Equipment ($/kW)	1000–1600	NA	NA	NA	NA	NA
	Maintenance ($/kWh)	0.006–0.019					
Wellhead gas	Equipment ($/kW)	NA	NA	NA	1550–2500	NA	NA
	Maintenance ($/kWh)				0.024–0.032		
Wood (forest residues)	Equipment ($/kW)	1250–2000	NA	NA	NA	NA	NA
	Maintenance ($/kWh)	0.008–0.023					
Wood waste	Equipment ($/kW)	1300–2100	NA	NA	NA	NA	NA
	Maintenance ($/kWh)	0.008–0.024					

[a] Including boiler costs.
[b] Including digester costs.
[c] Including gasifier costs.

Table 3 Comparisons of electric, thermal, and combined heat and power facilities

	Size (MW)	Fuel use (green ton/yr)	Capital cost (million $)	O&M[a] (million $)	Efficiency (%)
Electrical					
Utility plant	10–75	100,000–800,000	20–150	2–15	18–24
Industrial plant	2–25	10,000–150,000	4–50	0.5–5	20–25
School campus	NA	NA	NA	NA	NA
Commercial/ institutional	NA	NA	NA	NA	NA
Thermal					
Utility plant	14.6–29.3	20,000–40,000	10–20	2–4	50–70
Industrial plant	1.5–22.0	5,000–60,000	1.5–10	1–3	50–70
School campus	1.5–17.6	2,000–20,000	1.5–8	0.15–3	55–75
Commercial/ institutional	0.3–5.9	200–20,000	0.25–4	0.02–2	55–75
Combined heat and power (CHP)					
Utility plant	25 (73)[b]	275,000	50	5–10	60–80
Industrial plant	0.2–7 (2.9–4.4)	10,000–100,000	5–25	0.5–3	60–80
School campus	0.5–1 (2.9–4.4)	5,000–10,000	5–7.5	0.5–2	65–75
Commercial/ institutional	0.5–1 (2.9–7.3)	5,000	5	0.5–2	65–75

[a] Operating and maintenance.
[b] Sizes for the CHP facilities are a combination of electrical and thermal; the first figure is electrical and the figure in parentheses is thermal. 1 MW = 3.413 million Btu/h.
Source: From www.fpl.fs.fed.us/documents/techline/wood_biwnass_fas_energy.pdf (see Ref. 11).

program known as the Clean Development Mechanism. Joint Implementation is another mechanism built into the Kyoto Protocol that funds GHG reduction projects in countries with economies in transition, including the former Soviet-bloc nations.

The potential for the reduction in GHG emission in the United States has been modeled based on adherence to proposed national renewable energy standards. Fig. 7 illustrates that a 15% reduction in U.S. power plant carbon dioxide emissions could be achieved due to the displacement of traditional fossil fuels by renewable energy resources through 2020.

CONCLUSION

Opportunity fuels are those wastes or process byproducts that have significant energy content and could be used to

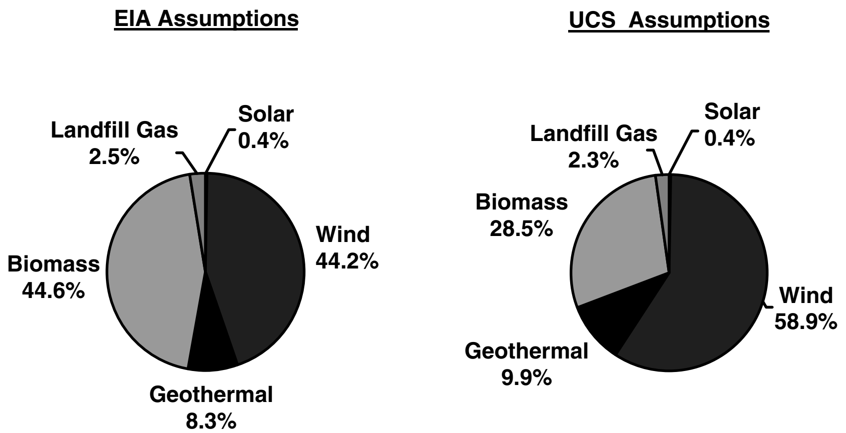


Fig. 5 Two predictions for the U.S. renewable generation mix in the year 2020.
Source: From Ref. 9.

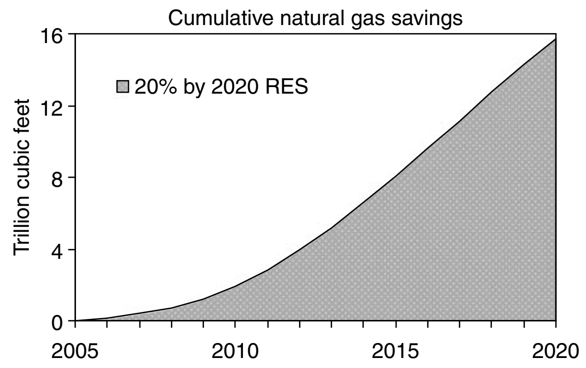


Fig. 6 Renewable energy conserves natural gas supply.
Source: From NREL Energy Analysis Seminar (see Ref. 10).

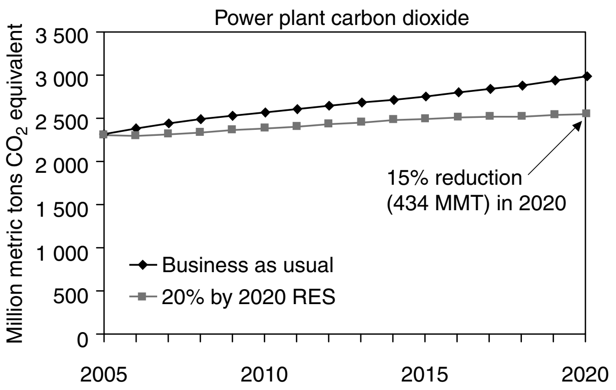


Fig. 7 Renewable energy reduces emissions and environmental compliance costs.
Source: From NREL Energy Analysis Seminar (see Ref. 10).

provide energy to generate electricity but have not traditionally been used for that purpose.

Relative to the combustion of traditional fossil fuels, burning or otherwise thermally processing opportunity and waste fuels offers the promise of reduced energy supply cost, with attendant environmental benefits including reductions in the amount of material going to landfills and a possible net reduction in carbon dioxide emissions.

Current interest in the utilization of these low-cost fuels for both thermal and electrical energy production is strong and is expected to increase due to the continuing upward trend in natural gas and oil prices. Further enhancing the appeal of prospective projects involving the use of these fuels and the technologies that exploit them is the potential for reduced environmental permitting time. Shorter project implementation times reduce costs and increase the rate of return on investment.

Glossary

Alternate fuel: Alternate fuel is a general term that can apply to any fuel not normally used by a process, boiler, or furnace, but that could be used either with or without appropriate changes to burner, fuel and fuel residue handling, or pollution control systems.

Co-firing: Co-firing is the combustion of a waste fuel with a traditional fossil fuel or fuels in a burner, boiler, or furnace—generally, as a low-cost substitute for another conventional fuel.

Industrial energy recovery: Industrial energy recovery is the reuse of energy exhausted in process waste streams or steam condensate, typically to heat reactants, feedwater, or combustion air.

Municipal waste: Municipal waste consists of a variety of materials (some combustible and some not), typically including trash; yard wastes; garbage; household waste (including paper, wood, and carpet waste); textile waste; plastic and other synthetic packaging materials; and minor amounts of metals, soil, and other noncombustibles.

Nonhazardous waste fuel: A nonhazardous waste fuel is a combustible organic material that is a byproduct of a manufacturing process; a waste packaging or shipping material; a component of municipal solid waste such as paper waste, yard waste, and household and office waste; and any other such materials other than those classified by a government regulatory authority as hazardous that are capable of being burned or co-fired in a boiler or furnace.

Opportunity fuel: An opportunity fuel is one that could be used economically as a source of energy to generate electric power but is not normally used for this purpose.

REFERENCES

1. Lemar, P. Jr. *Combined Heat and Power Market Potential for Opportunity Fuels*; Resource Dynamics Corporation: Vienna, VA, 2005.
2. Lemar, P. Jr. *Market Potential for Opportunity Fuels in DER/CHP Applications*, CHP Subcontractors Coordination Review Meeting, Washington, DC, April 14, 2005.
3. http://www.fwc.com/pow_services/parts/burners/Co_Firing_of_Biomass.cfm Foster Wheeler Corporation (accessed January 2006).
4. http://nett21.gec.jp/jsim%5F3/html/doc%5F458.html Takuma Company, Ltd (accessed November 2005).
5. http://www.eere.energy.gov/biomass/large_scale_gasification.html (accessed January 2006).
6. http://www.nyserda.org/programs/pdfs/digestergrantlist.pdf (accessed December 2005).
7. http://www.ext.nodak.edu/extpubs/ageng/structu/ae1015.htm (accessed December 2005).
8. http://www.eere.energy.gov/state_energy_program/ (accessed January 2006).
9. *Renewable Energy Can Help Ease Natural Gas Crunch*, UCS, March 2004.
10. Clemmer, S. *Renewable Energy Modeling Issues in the National Energy Modeling System*, NREL Energy Analysis Seminar, Washington, DC, December 9, 2004.
11. http://www.fpl.fs.fed.us/documnts/techline/wood_biomass_-for_energy.pdf (accessed January 2006).

BIBLIOGRAPHY

Relevant materials are too numerous to mention, however a few highlighted references are provided below.

1. Summary of the Proceedings of the International Conference on the Co-Utilization of Domestic Fuels, University of Florida Conference Center, Gainesville, FL, February 5–6, 2003.
2. Gielen, D.; Unander, F. *Alternative Fuels: An Energy Technology Perspective*, International Energy Agency Workshop on Technology Issues for the Oil and Gas Sector, Paris, France, January 13–14, 2005.
3. Tillman, D.A.; Johnson, D.; Miller, B.G. *Analyzing Opportunity Fuels for Firing in Coal-Fired Boilers*, 28th International Technical Conference on Coal Utilization and Fuel Systems, Clearwater, FL, March 10–13, 2003.

Waste Heat Recovery

Martin A. Mozzo, Jr.
M and A Associates Inc., Robbinsville, New Jersey, U.S.A.

Abstract
Many industrial, commercial, and institutional uses of energy result in excessive rates of waste heat rejection. Heat rejection is typically inherent in process uses; however, it may be utilized to meet other needs. Recovering and reusing rejected heat is known as waste heat recovery. Waste heat is usually recovered in the forms of steam, hot water, or hot air. The recovery medium is dependent on the quality of the waste stream, the potential use for the waste heat at the host, and the cleanliness of the waste stream. In this entry, the energy engineer is introduced to issues that should be considered in the economical and technical evaluation of waste heat recovery potential. These issues include: (1) the quality of the waste heat stream; (2) the calculation of the availability and applications of waste heat; and (3) the types of heat recovery equipment available.

INTRODUCTION

In many industrial and commercial energy applications, only a portion of the energy input is used in the process. The remainder of the useful energy is rejected to the environment. This rejected energy may potentially be recaptured as useful energy through waste heat recovery. Not all rejected energy can be recovered due to quality, usefulness in a host's load profile, and/or economical reasons that may make its recovery infeasible. This entry will serve as a guide to waste heat recovery in order to provide a framework for the energy engineer to develop waste heat recovery projects. It will discuss the concept of quality vs quantity, engineering concerns in waste heat recovery, sample calculations for waste heat recovery, and the types of waste heat recovery equipment that can be used.

THE CONCEPT OF QUALITY VS QUANTITY

While at first glance, a waste heat recovery project may appear feasible, this may not always be the case. The concept of quality vs quantity plays an important role. For the purposes of this entry, there are three classifications of waste heat. These are: (1) high-grade waste heat, generally 1000°F and above; (2) medium-grade waste heat, generally in the range of 400°F–1000°F; and (3) low-grade waste heat, generally below 400°F. Typically, the higher the grade of waste heat, the better the application for a successful and economical waste heat recovery project. This is referred to as the"quality"of the waste heat. It is better to have marginal amounts of high quality waste heat than large quantities of lower-grade waste heat.

Keywords: Waste heat recovery; Equipment; Heat conservation; Heat transfer.

There are numerous reasons why higher quality waste heat recovery provides a better application. One reason is that the waste heat stream has a much higher temperature than the recovery medium, resulting in higher temperature differences, and, thus, waste heat recovery equipment sizes can be smaller. This is a result of the higher heat transfer efficiency between the waste heat stream at a higher temperature to a lower temperature heat recovery medium. Second, the recovery medium can take on many forms, namely air, steam, or hot water, whichever is best suited for the host facility, whereas, with lower quality heat recovery, the form may be very limited, especially with low-grade waste heat. For example, if you have a waste heat stream that is at 400°F, steam is not a viable heat recovery medium, as the temperature difference will be too low. Finally, cost and savings become important issues in regard to the quality of the waste heat recovery system. Generally, the higher the quality of waste heat, the lower the capital costs for waste heat recovery equipment. This is because better heat transfer efficiencies will be realized with heat recovery equipment. Additionally, the higher the quality of the waste heat stream, the higher the savings usually are. Thus, with lower costs and higher savings, simple payback periods will be shortened, a result which is generally well-received by management.

A word of caution should be offered regarding the quality of waste heat. There are vendors who will reduce the grade of the available waste heat so that their equipment can be used in a heat recovery application. One such vendor manufactured only low temperature heat transfer equipment, which he was trying to sell as waste heat recovery equipment. His proposals included the dilution of high-grade waste heat streams by mixing the high temperature waste heat stream with ambient air in order to bring the temperature down to levels that his equipment could handle. His argument was that the total Btu content of the waste heat stream was not changed, just the temperature. Unfortunately, in reducing the temperature and quality of the waste heat stream, the size of the heat recovery equipment had to be increased to recover the waste energy, and the type of waste heat recovery medium was limited. This results in higher costs, lower savings, and longer-term payback periods.

There is one instance where the dilution of a waste heat stream may be justified. Steel is usually used in the manufacture of waste heat equipment, and there is an upper temperature limit for the waste stream where steel can be used. A substitution of other materials, for example, ceramics, can be made; however, the capital costs may become too prohibitive. In this instance, the dilution of the waste heat stream may be justified to bring the waste stream temperature down for the safe operation of the waste heat recovery equipment. It should, however, never be diluted simply to enable the use of a particular piece of heat recovery equipment. The primary reason is that by diluting the waste stream, energy savings may decrease and equipment costs will definitely increase; thus, payback periods will increase, perhaps significantly.

ENGINEERING CONSIDERATIONS

There are several engineering factors that must be evaluated when considering and designing a waste heat recovery system. In terms of steps, these are: (1) quantifying the waste heat stream; (2) determining the value of the waste heat stream; (3) evaluating the best form of heat recovery for the host facility; (4) determining the host site heat load profile; (5) determining the grade of waste heat; (6) determining the cleanliness and quality of the waste stream; and (7) selecting the proper waste heat recovery equipment by considering size, location, and maintainability.

The first step that should be executed is to quantify the waste heat stream by determining how many Btu/h are in the waste stream. The equation to calculate this is as follows:

$$Q = M \times \text{specific heat} \times \text{delta temp} \qquad (1)$$

where Q = total heat flow rate of waste stream in Btu/h; M = mass flow rate in Lb/h; specific heat = for air, 0.24 (Btu/Lb/°F); delta temp = ($T_{\text{upper}} - T_{\text{lower}}$) in °F.

The specific heat (Cp) changes as the temperature of the air rises. For example, at 1000°F, Cp is 0.26. Since this value is higher than that of Eq. 1, using the lower value, rather than correcting for temperature, provides a conservative calculation of the total heat rate in the waste stream.

The mass flow rate (M) is calculated as follows:

$$M = \rho \times V \qquad (2)$$

where M = mass flow rate in Lb/h; ρ = density of the waste stream in Lb/ft^3; V = volumetric flow rate in ft^3/h (in standard cubic feet).

Note that the value of Q is not the total amount of waste heat that will be recovered, but, rather, the total amount of waste heat that is ideally available for recovery. Not all of this waste heat will be recovered, or even can be recovered. The total amount that will be recovered will be determined by numerous other factors, such as the cleanliness of the waste stream and the form of recovery (i.e., high pressure, superheated steam, saturated steam, or hot water). This step is necessary in determining if there are sufficient volumes available for waste heat recovery.

One common mistake committed when quantifying the amount of waste heat available is assuming values for flow and/or temperatures without making measurements. Too often, actual conditions vary from assumed conditions, a variation that can often cause disastrous results in a waste heat recovery project. The cost of making actual field measurements is a small price to pay to obtain reliable data.

After the quantity of the waste stream is correctly determined, step two is to find the dollar value of the waste heat stream to determine how much capital cost a potential project can bear, or if it even makes economical sense. In all waste heat recovery projects, the heat recovered will displace a medium, such as steam, which would have to be generated using another piece of equipment, such as a boiler. This equipment likewise has a related efficiency, and the heat output is always less than the heat input. To determine the dollar value of the waste heat stream, use Eqs. 3 and 4 below:

$$\text{Value} = Q \times \text{unit cost} \qquad (3)$$

where value = the dollar value of the waste heat stream, per hour; Q = total heat flow rate of waste stream in Btu/h as calculated from Eq. 1; unit cost = unit cost of the waste stream in dollars/Btu.

$$\text{Unit cost} = \frac{\text{fuel cost}}{\text{efficiency}} \qquad (4)$$

where unit cost = dollars/Btu; fuel cost = cost for fuel displaced in dollars/Btu; efficiency = efficiency of unused equipment. For example, a steam boiler @ 75%.

The third through fifth steps tend to overlap and be interrelated, and are discussed together. In order for waste heat recovery to be acceptable to a host facility, its use must be consistent with current energy usage at the facility. To best apply waste heat recovery, the engineer needs to examine current energy usage patterns, heat loads for the site, and the quality of the waste heat that is available. The load factor of the waste heat recovery stream is the first thing to determine. For example, the waste heat recovery stream may be exhaust air flows from a process furnace that is periodic in nature, rather than a continuous operation. In this case, the best potential would be to return the waste heat to the process in some manner that allows the usage to follow the waste heat stream generation. Examples include utilization in a drying operation or as pre-heated combustion air for the process furnace. In another situation, the process furnace may run continuously and be of a high enough grade to generate steam; however, if the host facility only uses steam for heating in the winter, the loads do not match and, thus, steam is not a good recovery medium. On the other hand, there may be a need for large quantities of hot water in the facility, and the loads may match up. A thorough evaluation of both the source and operation of the waste heat and potential uses to recover the waste heat must be performed.

The sixth step, determining the cleanliness and quality of the waste stream, is important. If the waste heat stream is the product of the combustion of a natural gas-fired or fuel oil-fired operation, then the gases should be relatively safe for waste heat recovery. The important issue here would be the condensation of acidic liquids in the products of combustion gas stream, especially if fuel oil is used. This condensation would happen if the temperature of the waste stream is allowed to drop to low temperatures, typically below 300°F. The waste stream could also contain a burn-off from the product itself, which could condense on the waste heat recovery equipment and cause blockages. Here, it is best to maintain the exit temperature significantly above the condensation temperature to avoid the formation of acids or other potential hazards from the waste stream.

If the waste heat stream contains the products of combustion and the products of the process, great care must be exercised when utilizng the waste heat. The selection and design of waste heat recovery equipment could lower the waste heat exhaust temperature below acceptable levels, resulting in condensation or blockage problems with the heat recovery equipment. When unsure of the condensation temperatures and the effect on the heat recovery equipment, simple tests must be conducted to obtain this information.

Some actual examples wherein the waste heat stream was dirty and created problems follow. The first was an air-to-air heat recuperator used in a process operation. The products of combustion gases were dirty, creating a buildup on the heat exchanger. A soot blower was used to remove the buildup on a periodic basis; however, this was not totally successful. During actual heat recovery operations, the soot blowers could not remove the buildup adequately. A solution to this problem occurred accidentally. The operation was a 24 h/five days a week, and by mistake one weekend, the soot blower was left on, while the waste heat recovery equipment was shut off. Because the buildup and the heat exchanger were composed of different materials, the heat exchanger and the buildup thermal contraction rates were different. As a result, the soot blower was able to break up the buildup on the heat exchanger surfaces. This method was then used to clean the heat exchanger each week. The efficiency of the heat exchanger did decrease somewhat during the week, though not significantly. During the weekend, the buildup was blasted off the heat exchanger surface, thus increasing the heat exchanger's efficiency at the beginning of another week.

The second example did not turn out as well. In this scenario, the quality of the waste heat stream was low, and the products of combustion, combined with product burn-off, created a dirty waste stream. As a result, a serious problem quickly developed. With the waste heat recovery system in operation, gas products in the waste heat stream condensed on the heat exchanger surfaces, and within a very short period of time, the heat exchanger was completely blocked. The only corrective action was to remove the waste heat recovery equipment from service and steam clean it, a very costly operation that eventually resulted in the failure of the project.

The seventh step in the design of a heat recovery system is the selection of the waste heat recovery equipment. Descriptions of several different types of equipment are given in another entry in this encyclopedia. Selection of the proper piece of equipment will be based on the quality of available waste heat; the recovery fluid that can be generated, i.e., steam, hot water, or air; the location of the equipment; and the maintenance capabilities of the host facility. While the first two criteria should be obvious, the last two are sometimes neglected.

The location where the equipment is to be placed is important. It obviously should be close to the waste heat stream, yet should not interfere with the other equipment involved in the generation of waste heat, for example, a process furnace. The equipment also must not be squeezed into an area where maintainability is an issue. Sufficient access for maintenance and overhaul must be included since these pieces of equipment will require periodic maintenance. Accessibility to perform maintenance is not just desirable, but mandatory.

SAMPLE CALCULATIONS

This section describes some sample calculations that are used in the design of waste heat recovery systems. The first step is to calculate the amount of waste heat that is available for recovery. This is done using Eq. 1.

Example—Waste heat is available in the form of hot products of combustion from a natural gas-fired furnace. Assume that there are no contaminants in the waste heat stream. Measurements indicate that there are 14,000 available actual cubic feet per minute (ACFM) flowing from a process at 1000°F. The actual volumetric flow must first be converted to standard cubic feet per minute (SCFM) at 60°F, using the following equation:

$$\text{SCFM} = \text{ACFM} \times (T_{\text{absolute}} + 60)/(T_{\text{absolute}} + T_{\text{actual}}) \quad (5)$$

where SCFM = standard cubic feet per minute; ACFM = actual cubic feet per minute; $T_{\text{absolute}} = 460°F$; $T_{\text{actual}} =$ actual gas temperature.

Using the data available here,

$$\text{SCFM} = 14{,}000 \times (460 + 60)/(460 + 1000) = 4{,}986 \text{ ft}^3/\text{min}$$

Since the product of combustion is essentially air, a good approximation of the density of the gases is similar to that of air at standard conditions, or $\rho = 0.074$ Lb/ft^3.

Using Eq. 2, we can calculate the mass flow rate for the waste stream as follows:

$$M = \rho \times V$$
$$= 0.074 \text{ Lb/ft}^3 \times 4{,}986 \text{ ft}^3/\text{min} \times 60 \text{ min/h}$$
$$= 22{,}139 \text{ Lb/h}$$

The next step is to calculate how much waste heat is available for transfer to the heat recovery system. The upper temperature has been given as 1000°F. The lower temperature will be determined by the medium used for heat recovery, that is, steam, hot water, or air, as well as the cleanliness and condensation temperature of the waste stream.

To continue the sample calculations, we will use three different applications of heat recovery mediums: 125 PSIG steam, the hot water required at 350°F, and pre-heated combustion air at whatever final temperature is available. Each of these applications is analyzed separately.

The first heat recovery application medium is saturated steam at 125 PSIG (140 PSIA). From steam tables, we find the enthalpy for saturated liquid (h_f) to be 324.82 Btu/Lb. and for saturated vapor (h_g) 1193.0 Btu/Lb. Thus, we need 868.18 Btu for every pound of steam (1193.0 − 324.82). The temperature of steam at this pressure is 353°F. For this temperature of steam, we can take the waste stream down to approximately 400°F without the risk of condensation (since condensation usually does not occur above about 325°F). Using Eq. 1, we can calculate the available waste heat as follows:

$$Q = 22{,}139 \text{ Lb/h} \times 0.24 \text{ Btu/Lb/°F} \times (1000 - 400)$$
$$= 3{,}188{,}041 \text{ Btu/h} = 3.188041 \text{ MMBtu/h.}$$

This waste heat stream's value can be calculated using Eqs. 3 and 4. Assuming we have natural gas priced at \$8.00/MMBtu in a steam boiler with an efficiency of 75%, the value becomes:

$$\text{Unit cost} = \$8.00/\text{MMBtu}/0.75 = \$10.67/\text{MMBtu}$$

$$\text{Value} = \$10.67/\text{MMBtu} \times 3.188{,}042 \text{ MMBtu/h}$$
$$= \$34.00/\text{h}$$

To obtain steam mass flow, we must use Eq. 6 below:

$$M_s = \frac{Q}{(h_g - h_f)} = \frac{3{,}188{,}041 \text{ Btu/h}}{(1193.0 - 324.82)\text{Btu/Lb}} = 3{,}672 \text{ Lb/h.} \tag{6}$$

The second heat recovery application medium will be hot water leaving the heat recovery equipment at 250°F, using the same waste stream. Water will be supplied to the heat recovery equipment at 200°F. With the temperatures involved, we can take the waste stream down to approximately 300°F, provided this low temperature does not cause any condensation issues. For the purposes of this example, we have determined that condensation occurs below 325°F. We will use 350°F as our lower limit. Using Eq. 1, we can calculate the available heat as:

$$Q = 22{,}139 \text{ Lb/h} \times 0.24 \text{ Btu/Lb/°F} \times (1000 - 350)$$
$$= 3{,}453{,}711 \text{ Btu/h.}$$

In this case, we need to determine the mass flow rate, expressed as gallons per minute (GPM), of hot water we can heat from 200 to 250°F. Eq. 7 below provides the methodology:

$$Q' = 500 \times V \times (T_{\text{upper}} - T_{\text{lower}}) \tag{7}$$

where Q' = total heat recovered in Btu/h; V = volumetric flow rate of water in GPM. 500 is the constant used to convert GPM to Lb of water per hour.

To calculate the volume of water required for this system, we calculate:

$$V = \frac{Q'}{500 \times (T_{\text{upper}} - T_{\text{lower}})} = \frac{3{,}453{,}711 \text{ Btu/h}}{500 \times (250 - 200)} = 138 \text{ GPM}$$

Thus, our waste heat recovery system will generate approximately 138 GPM of hot water raised from 200 to 250°F.

The third heat recovery application medium is an air-to-air recuperation system. The waste stream is the same as in the previous example, providing 3,453,711 Btu/h (assuming that 350°F will be the lower limit that we can go to without condensation issues). The recovery medium will be air applied as pre-heated combustion air with an input temperature of 75°F. In this system, we need to provide a set volume of combustion air to the waste heat recovery equipment, so we are interested in finding the final temperature of the pre-heated combustion air.

Using Eq. 1:

$$Q = M \times 0.24 \times (T_{upper} - T_{lower})$$

$$3,453,711 \text{ Btu/h} = 50,000 \text{ Lb/h} \times 0.24 \text{ Btu/Lb} -°F \times (T_{upper} - 75)$$

Solving for T_{upper} we get 362.8°F

HEAT RECOVERY EQUIPMENT

There is a variety of equipment manufactured and/or sold as heat recovery equipment. The following is a list of some more common types of equipment.

Waste heat steam recovery—This piece of equipment can be either a water-tube or fire-tube boiler that uses the hot waste heat gas stream to heat boiler feed water to generate either low-pressure or high-pressure steam. Typically, this piece of equipment is called a heat recovery steam generator (HRSG). The unit can be a heat recovery-specific piece of equipment; however, some boiler manufacturers will sell their boilers without a burner package and call them HRSGs. This type of boiler can lose some of its effectiveness as a steam generator; however, it still is an acceptable piece of equipment. The reason for the decrease in recovery effectiveness is solely based on the possibility that the waste heat stream temperature will be somewhat lower than the burner flame temperature.

Recuperator—This piece of equipment is generally an air-to-air heat exchanger that transfers heat from the waste heat gas stream to air on the recovery side of the heat exchanger. This air can be used as pre-heated combustion air, or host make-up air in the facility.

Shell and tube heat exchanger—This piece of equipment consists of a bundle of tubes within a steel shell. It is usually used for water-to-water heat recovery. One of its uses may be to recover heat from a boiler condensate blow down on the shell side to heat up boiler feed water on the tube side.

Fin-tube heat exchanger—This type of heat exchanger uses air (usually the waste heat stream) blowing across finned coils that contain water. A typical application for this type of heat exchanger is using boiler products of combustion gases to preheat boiler feed water. Plugging of the finned coils could be a problem if the fins are closely spaced, the exhaust gases are dirty, or the condensation temperature (of the gases) is approached.

Plate and frame heat exchanger—This piece of equipment consists of two frames sandwiching thin plates. It is usually used with fluids of relatively low temperature, which flow through alternate plates. Since the plates are thin, heat transfer is usually good; however, both fluid streams need to be relatively clean or the exchanger will plug up.

Heat wheels—These are typically used in low temperature applications, such as the exhaust of environmental (space) air coupled with the introduction of outside air. The two ducts, outside and exhaust air, must be adjacent to one another so that the wheel turns through both ducts. The wheel will turn at low revolutions per minute (RPM) collecting the waste heat or cooling, depending on what season it is, and exchange the heat or cooling to the outside air being introduced to the facility.

Heat pipes—This equipment consists of a pipe heat exchanger with the interior containing a coolant. The coolant is alternately vaporizing and condensing between an exhaust and outside air stream, exchanging cooling and heating between the two air streams. The exhaust air and outside air ducts must reside together in the same manner as a Heat Transfer Wheel.

Run-around coils—These are similar to heat pipes in operation. They cool or heat exhaust air and outside air streams. Their typical construction is a finned water coil with air blown across the coil. The advantage of the run-around coil is that the exhaust air and outside air ducts can be physically separated. The disadvantage is that a pump, with pumping power, is required. Additionally, if the coils are subject to freezing conditions, then a water/glycol solution must be used. This solution will decrease the effectiveness of the heat transfer.

SUMMARY

The purpose of this entry has been to provide the energy engineer with some general guidelines to applying waste heat recovery. There are numerous applications in the industrial, institutional, and commercial sectors where waste heat recovery can be used cost effectively. The best applications are situations wherein the waste heat stream is of high quality, produced for many hours of the year, preferably 24/7, with a heat load that matches the waste heat availability. Care must be taken by the engineer to ensure that the waste heat stream will not block or destroy the waste heat recovery equipment.

There are also numerous pieces of equipment that can be used for waste heat recovery. These include steam boilers, hot water boilers, recuperators, coil heat exchangers, shell and tube heat exchangers, plate and frame heat exchangers, heat transfer wheels, heat pipes, and run-around coil systems.

BIBLIOGRAPHY

1. Thumann Albert, P.E.; C.E.M; Metha, D.P. *Handbook of Energy Engineering*; The Fairmont Press, Inc.: Lilburn, GA, 1997.
2. Turner, W.C. *Energy Management Handbook*, 3rd Ed.; The Fairmont Press, Inc.: Lilburn, GA, 1997.

Waste Heat Recovery Applications: Absorption Heat Pumps

Keith E. Herold
Department of Mechanical Engineering, University of Maryland, College Park, Maryland, U.S.A.

Abstract
Absorption refrigeration/heat pump technology is a heat-driven system for transferring heat from a low temperature to a high temperature. The major current application is in the building cooling market where such machines provide a natural-gas-fired cooling option which is particularly popular in markets where electric costs are high in comparison with natural gas costs. Another set of applications exists where the absorption machines are fired by waste heat, typically acting as a bottoming cycle for an integrated energy system. Examples include turbine inlet air cooling, engine cooling jacket heat recovery, and energy recovery in chemical plants. At current energy prices, waste heat has little value and is often discarded even though it has thermodynamic potential. Projections of the depletion of fossil fuel reserves and projections of the environmental implications of wasteful burning of fossil fuels seem to point toward a future where the value of waste heat would increase relative to other commodities. In such a future, it is expected that waste-heat-driven absorption technology would become more widespread as a natural bottoming cycle for many high temperature processes. This entry summarizes the potential for the use of absorption cycles for waste heat recovery.

INTRODUCTION

Absorption chillers and heat pumps are machines that transfer heat from a low temperature to a high temperature.[1–4] As we know from the second law of thermodynamics, such a process requires an energy input. In the case of absorption technology, the energy input is in the form of a high temperature heat transfer into the system that drives the process. Absorption machines typically also require an electricity input to power the pumps and the control system, but the primary energy input is heat transfer. As such, absorption machines have an important role in waste heat utilization. In the current economic climate, characterized by relatively low energy costs, waste heat utilization does not get much attention. However, a trend toward more conservative use of energy appears to be gaining momentum based on economic and environmental concerns. This trend is expected to lead to resurgence in the demand for waste-heat-fired absorption technology.

THERMAL DESIGN FUNDAMENTALS

The ability of a heat transfer to do work (including providing refrigeration) depends on its temperature according to the well-known Carnot law

Keywords: Absorption chiller; Heat pump; Waste heat; Cogeneration; Combined heat and power (CHP); Turbine inlet air cooling.

$$W_{\max} = Q\left(1 - \frac{T}{T_o}\right)$$

where $W_{\max}$ is the maximum possible work output, Q is the heat transfer, T_o is the ambient temperature and T is the temperature at which the heat transfer occurs. Note that the temperatures must be expressed in absolute temperature units (e.g., K). For heat transfer processes that occur over a range of temperatures, the temperature T can be replaced by an average temperature to give a reasonable estimate. The fact that the same amount of heat transfer can do more work if it occurs at a higher temperature comes directly from the second law of thermodynamics and is one of the keystones of thermal design. One of the corollaries is that it is important to provide temperature matching between a heat source and a process that requires heat, so as to avoid wasting the potential to do work.

Combustion temperatures can be quite high, easily above 1000°C. However, the strength of the metals used in the construction of energy systems usually limits high temperatures to values of less than about 500°C. This value depends on the materials and on the mechanical stresses that the system experiences during operation. Temperature matching at the high temperature end would argue for an approach closer to combustion temperatures, but the materials issues are significant, intertwining both economics and materials science.

One of the requirements of the second law of thermodynamics is that all thermal power cycles must reject a portion of the high temperature heat input as heat at the low temperature of the cycle. Another way to say this is that the heat transfer input cannot be converted completely into work. The energy input as heat is only partially converted to work and the remainder of the energy must be rejected as low temperature heat. This low temperature heat is often referred to as waste heat (in general, any heat that is not used in a process can be called waste heat). Depending on the thermal efficiency of the power cycle, the rejected energy can be significant. For example, if the power cycle is 33% efficient, then 67% of the input energy is rejected as heat. This waste heat is a resource, but it is important to realize that its value is in its ability to do work, which is influenced both by its temperature and by the energy content. If the waste heat is at ambient temperature, then it has zero value even though it may be a significant energy flow.

BOTTOMING AND TOPPING CYCLES

In the design of a thermal power cycle, the thermal efficiency can, in general, be maximized by introducing the input heat at the highest possible temperature and rejecting heat at the lowest possible temperature. However, design details often restrict the ability to do this (e.g., materials issues at the high temperature end) and the waste heat is often rejected at a temperature sufficiently far above ambient so as to be useful. Examples include: the waste heat from an automobile engine where the exhaust gases and the jacket cooling fluid are both at temperatures well above ambient; the waste heat from a combustion turbine.

In the practice of thermal design, the idea of using a heat transfer stream twice (or more times) as it cascades down from high temperature to ambient, is often discussed in terms of topping and bottoming cycles. When the waste heat from a power cycle is used to drive an absorption chiller, the power cycle is called the topping cycle and the absorption cycle is the bottoming cycle. Both cycles are driven by the same heat transfer as it cascades down in temperature. Although in theory, an absorption cycle could be inserted as a topping cycle, materials considerations limit the range of temperatures for the most common working fluids limiting the capability of an absorption machine for use as a bottoming cycle. The absorption cycle must also reject heat, but if the temperature of the rejected heat is close to ambient, than it has little value even if a large energy flow rate is present.

The bottom line for practical waste heat utilization is that the temperature of the waste heat must be sufficiently different from the ambient temperature to make the stream look attractive. Beyond that, the particular technology considered for utilizing the waste heat must be well matched to the waste heat and to the load. For example, a standard (so-called single effect) absorption cycle requires a heat input of around 120°C and a cooling tower to reject the heat. The machine provides chilled water as the output. If all three heat sources/sinks exist in a given application, then it is a good match for the absorption machine.

ABSORPTION HEAT PUMP CONFIGURATIONS

There are a number of useful ways to categorize absorption chillers and heat pumps. These include:

- Working fluid systems
- Heat source and sink configurations

In a sense, the heat source and sink configuration is more fundamental than the working fluid system, although in a practical sense these categories overlap and influence each other.

Heat Source and Heat Sink Configurations

Many heat source and sink configurations are possible with absorption technology if simple cycles are combined into more complex combinations. There are many publications about such cycles.[5–7] However, as an introduction to the subject, only the simplest cycles are considered here. For this purpose, it is assumed that the absorption machine interacts with three external heat sources or sinks at three distinct temperature levels.

Historically, absorption cycles are categorized as either Type I or Type II as shown in Figs. 1 and 2. Type I refers to the configuration found in building

Was–Wat

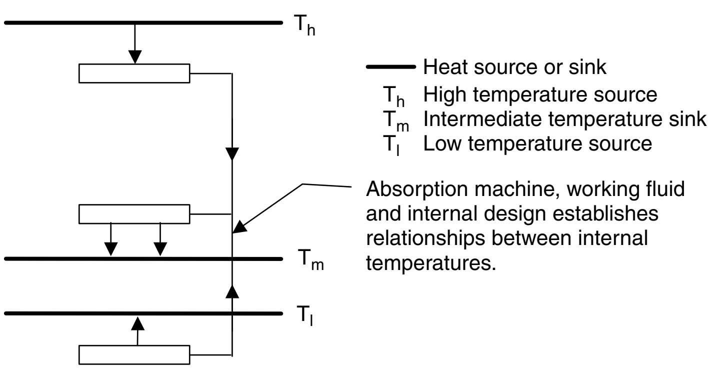


Fig. 1 Type I absorption heat pump. Acts as a chiller (refrigerator) if T_m is the ambient temperature sink. Acts as a heat pump if T_l is the ambient temperature.

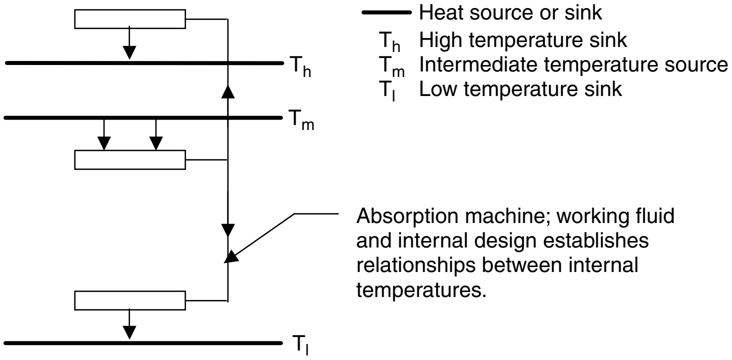


Fig. 2 Type II absorption heat pump. Also called a heat transformer or a temperature boosting heat pump.

air-conditioning applications and is the most well-known configuration. Although a Type I machine can act as either a chiller or a heat pump, the chiller application is, by far, more common due to the lack of competing technologies for the heat-driven chiller market. For a building's chilling application, the absorption machine will reject heat to ambient at the intermediate temperature (typically via a cooling tower). However, the same machine can be used for heating purposes by connecting the lowest temperature to the ambient. For both of these Type I configurations, the heat flow directions are the same as shown in Fig. 1. It is noted that the heat transfer out of the system at the intermediate temperature is equal to the sum of the heat transfers into the system at the high and low temperatures. Also, the maximum rate of refrigeration/heat pumping is limited by the second law of thermodynamics and depends on the temperatures of the three heat sources and sinks.

Type II heat pumps, as shown in Fig. 2, have been built and run as research machines but are not commercially available (because of the low current value of waste heat). In comparison with the Type I configuration, all of the heat transfers occur in the opposite directions in a Type II machine. The heat is input to the machine at the intermediate temperature, and a portion of the input energy is pumped out of the system at the highest temperature. The remaining energy is rejected to ambient at the lowest temperature. The purpose of a Type II heat pump is to upgrade the temperature of a heat transfer. An example application might be in a chemical plant where a component must be cooled at 100°C and elsewhere in the plant a component requires heat at 120°C. In such a case, the plant owner could burn some fuel to increase the temperature of the available heat by 20°C but a Type II heat pump provides an option that supplies the needed heat without requiring additional fuel. As was noted for the Type I heat pump, it is noted here that the sum of the energy out at the high and low temperatures must equal the input energy. Furthermore, the second law of thermodynamics puts constraints on the portion of the input heat that can be upgraded, which depends on the temperatures of the three sources and sinks.

Absorption Working Fluids

Absorption working fluids consist of a refrigerant and an absorbent. In general, the absorbent has a much higher boiling point than the refrigerant so that the refrigerant can be separated readily from the absorbent by heating. Over the years of absorption system development (which goes back to the 1800s), many working fluid pairs have been tried. A large number of thermophysical properties interact in a working absorption chiller/heat pump, and the performance of a particular working fluid pair can be derailed by many properties. Thus, the process of selecting working fluids is bound by many constraints. Although many fluid pairs can be made to work, there are only two pairs that have emerged in commercial systems. These two are: water/ammonia and lithium bromide/water.[8–11] In each of these pairs, the absorbent is listed first, followed by the refrigerant. Note that the water plays the role of absorbent in one case and refrigerant in the other.

Both water and ammonia make good refrigerants because they each have a high latent heat, resulting in a low refrigerant flow rate for a given heat transfer rate. Although water has twice the latent heat of ammonia, water suffers from two major drawbacks which limit its use: water's low vapor pressure, resulting in sub-atmospheric (vacuum) internal pressures and in-leakage of air; and the freezing point of water limits its use to temperatures above 0°C. On the other hand, the vapor pressure of ammonia is, in some sense, too high such that ammonia is restricted to applications with temperatures below 150°C to avoid high pressure on vessel walls. Many other refrigerants have been considered for use in absorption cycles, including the refrigerants used in vapor compression refrigeration, but water and ammonia are preferred because they are paired with effective absorbents.

A key property of an effective absorbent is that it must have a high affinity for the refrigerant (in other words, there needs to be a molecular attraction that makes mixing occur readily). Other beneficial properties include low vapor pressure and maintenance of a liquid phase over the operating temperature range of interest. The water/ ammonia pair remains liquid over the full concentration range and over a wide range of useful temperatures but the vapor pressure of water is sufficiently high so as to require special design features (i.e., a rectifier) to remove water vapor from the refrigerant after it is boiled out of solution. This extra component must be cooled resulting in a reduction in thermal performance for a water/ammonia machine. Lithium bromide is a good absorbent for water and has a very low vapor pressure, but it has limited solubility leading to crystallization (solid hydrate formation) that tends to plug up flow passages under conditions of high concentration. Crystallization can be avoided by maintaining low heat rejection temperatures; this implies that lithium bromide/water machines must operate with a cooling tower.

Lithium bromide/water systems utilize a surfactant as a heat and mass transfer enhancement factor to enhance internal transfer processes.[12–14]

Corrosion is an issue for both of the common working fluid pairs. In most cases, corrosion inhibitors are found to effectively limit corrosion to acceptable levels.

HOW DOES ABSORPTION REFRIGERATION WORK?

The easiest way to understand absorption refrigeration technology is to first understand vapor compression technology as shown schematically in Fig. 3. A vapor compression cycle consists of just four components: a compressor, a condenser, a throttle, and an evaporator. Mechanical energy is input to the compressor to raise the pressure of the refrigerant vapor, allowing it to condense into liquid in the condenser (accompanied by heat rejection). The liquid then passes through a throttle which lowers the pressure. A throttle is just a flow restriction, which can take the form of an orifice, a valve, or a capillary tube. Due to the drop in pressure across the throttle, a portion of the liquid evaporates and this cools the remaining liquid. The liquid then enters the evaporator where it absorbs heat from the outside of the device while it evaporates at low temperature. The cooling effect comes from the low pressure maintained in the evaporator by the compressor.

A similar schematic of an absorption refrigeration machine is shown in Fig. 4. Similarities to the vapor compression refrigeration machine can be seen on the left side including the condenser, throttle, and evaporator. The difference in the two technologies comes in how the compression process is accomplished. In the absorption machine, compression is accomplished via a thermal compressor consisting of the absorber, a solution heat exchanger, and the desorber (sometimes called the generator). The compression process goes as follows. First the vapor from the evaporator is absorbed into the liquid absorbent accompanied by heat rejection. Then the refrigerant-laden absorbent is pumped up to the higher pressure level in the desorber by a liquid pump. By the application of heat to the desorber, the refrigerant is then boiled out of the absorbent and sent to the condenser. The liquid absorbent leaving the desorber is hot and is brought into regenerative heat transfer contact with the solution leaving the absorber to minimize the heat requirements in both the absorber and the desorber.

One of the features of the thermal compressor is that the increase in pressure is accomplished while the refrigerant is in its liquid phase. In contrast to the compression of a vapor, the compression of a liquid requires very little energy because the small volume changes that the liquid experiences do not store much energy. This implies that the electrical energy input required by an absorption machine is minimal. Another aspect of the thermal

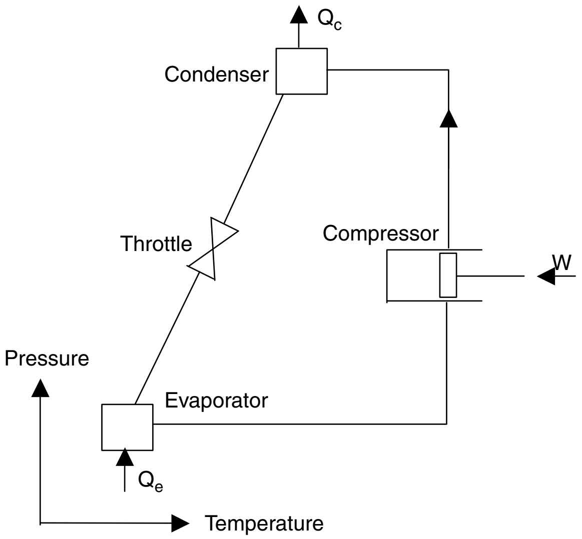


Fig. 3 Schematic of a vapor compression refrigeration/heat pump cycle.

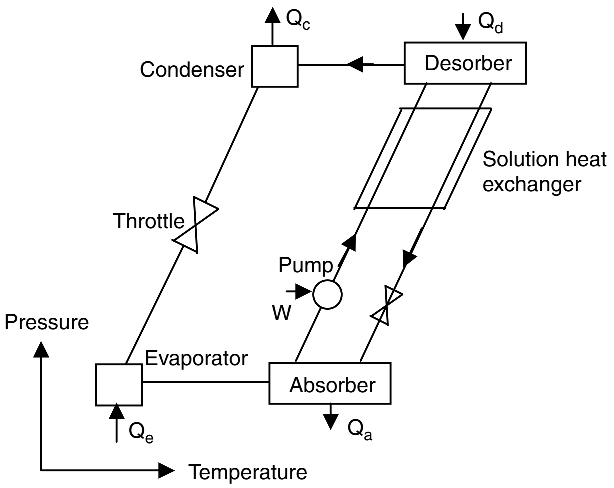


Fig. 4 Schematic of an absorption refrigeration/heat pump cycle.

compressor is the fact that the energy that must be rejected by the absorber is similar in magnitude to the energy that must be supplied to the desorber. Furthermore, this energy is somewhat greater than (on the order of 20% greater) the latent heat of the refrigerant due to the energy interactions between the absorbent and the refrigerant (associated with the affinity of the absorbent for the refrigerant; i.e., the heat of mixing). The presence of the heat of mixing further limits the thermal performance of an absorption machine.

ABSORPTION MACHINE PERFORMANCE

As with other refrigeration technologies, the thermal performance of an absorption machine is defined in terms of a coefficient of performance as

$$\eta_{I} = \frac{Q_e}{Q_d} \qquad \eta_{II} = \frac{Q_a}{Q_d + Q_e}$$

$$\eta_{vc} = \frac{Q_e}{W}$$

where Q_e is the evaporator heat (i.e., the refrigeration effect), Q_a is the absorber heat, Q_c is the condenser heat, Q_d is the desorber heat (i.e., the high temperature heat transfer), and W is the compressor work input. Typical coefficient of performance values are given in Table 1. The value listed in parentheses for the double effect machine is a gas-fired value that is lower than the machine value due to gas burner losses.

The range of values shown for the single effect, Type I machine represent differences in performance between the two standard working fluid pairs with water/ammonia having lower performance due to the energy penalty of the rectifier that is needed to strip water out of the refrigerant after the desorber. The values shown for the double effect machine are only lithium bromide/water values since the temperatures for double effect make water/ammonia less practical. The Type II value is a value obtained by the author on a research machine (unpublished).

The value given for the vapor compression technology represents a typical value for an application comparable to the absorption chiller applications. It is evident that the vapor compression technology has a high coefficient of performance but it must be understood that this is a somewhat biased comparison because the Carnot factor for the work input has a value of 1. A fairer comparison results if a power cycle efficiency factor is included in the calculations so that both technologies are compared based on a fuel input. If this is done with a power cycle efficiency of 33%, then the coefficient of performance of the vapor compression machine is reduced to 1.67. Although this comparison is more favorable to the absorption technologies, the conclusion is that vapor compression technologies, combined with high-efficiency electricity generation, are hard to beat for refrigeration. However, for the waste heat applications that are the focus of this article, the comparison to vapor compression technology is not relevant. If the input heat for the absorption machine is waste heat, then it implies that any output from the absorption machine is over and above what would otherwise have been produced.

Table 1 Typical coefficient of performance values for various absorption technologies, including vapor compression technology for comparison

Cycle configuration	η
Type I	
Single effect	0.5–0.72
Double effect	1.3 (1.1)
Type II	
Single effect	0.3
Vapor compression	
Typical	5

BUILDING COOLING APPLICATIONS

The use of absorption chillers for building cooling applications is widespread, with a market share of only a few percent in the U.S., but up to 50% in Japan, China, and South Korea. The primary technology that is utilized in these applications is the direct-fired, double-effect lithium bromide/water chiller. The primary driver in the building cooling market is the ability to provide building cooling with a gas or oil input. All of these markets are characterized by relatively high electricity costs in comparison to fuel costs. In the U.S., absorption chillers have significant market share in markets such as New York City where the electric infrastructure is expensive to upgrade but where natural gas is available. In Japan, government policy requires gas-fired cooling in an attempt to control electricity growth. In China and Korea, urban growth outpaced power plant construction leading to limited availability of electric power in some cities.

Building cooling using direct-fired absorption chillers makes sense economically in these locations but makes little sense thermodynamically. In particular, much of the potential of the fuel is lost when it is burned and then utilized at such low temperatures. The combustion process occurs at temperatures approaching 1000°C, while the heat input to the absorption machine is needed at around 250°C. Thus, there is a thermodynamic opportunity to extract more useful work from the fuel than is obtained when it is used for direct firing of an absorption chiller. In particular, it would be possible to run a topping cycle (e.g., a power cycle) that would make use of the high temperature combustion heat and which would be mated with the absorption chiller so that both machines would use the same energy input as it cascaded down in temperature.

This is not a new idea, but it is not widely used today due to various economic realities (i.e., small-scale power generation is difficult to integrate into a grid-connected electricity network, and small power generating equipment is less efficient and costly to maintain).

Since absorption technology has a unique role in energy integration, it is expected that the application mix for absorption chillers will shift over time from the current mix dominated by direct-fired machines to a future application mix dominated by waste heat fired machines.

TURBINE INLET AIR COOLING

Combustion turbines, fired by natural gas, are widely used for electric power generation due to a range of features including:

- Modular design
- Low initial cost compared to coal-fired power
- Relative ease of power modulation

These machines are modified gas turbines and operate at high rotational speeds, transferring the energy released from combustion into rotating power to drive a generator. These systems ingest ambient air at their compressor inlet which is then fed to the combustion chamber. After combustion, the hot gases are expanded through a power generating section. The power output is strongly dependent on the mass of air ingested which is a function of the turbine's rotational speed and the density of the air. The performance of the turbine is maximized when it operates at its maximum speed. However, because the air density varies with ambient conditions (temperature and pressure), the power output also varies. This is problematic in many applications because the electric demand often peaks when the ambient temperature is high (when electric usage is dominated by electric-drive cooling systems). Thus, these combustion turbines lose capacity when it is most needed.

One option for leveling the power output is to provide inlet air cooling.[15] This can be done in a number of ways, including evaporative cooling (useful in dry climates where water is available), vapor compression cooling (which can be electric- or steam-driven), or absorption cooling.[16,17] Absorption cooling is a good match in applications where the waste heat from the turbine is not otherwise used. When waste heat is available free, the only costs are the capital cost of the absorption chiller and the additional pressure drop due to the cooling coil in the inlet air stream. Turbine inlet air cooling by absorption chillers can be implemented with chillers based on either water/ammonia or lithium bromide/water working fluids with the latter being more prevalent due to the availability of standard chiller packages designed for this application.

One interesting aspect of the inlet air cooling application is that the thermal efficiency of the gas turbine system is not as strongly influenced by the inlet air temperature as is the system capacity (i.e., power output). One of the key material design issues in such a system is the maximum temperature that the turbine blades experience. On the operational side, this maximum temperature is governed largely by the ratio of the fuel and air flow rates. Thus, when the ambient temperature is low, the reduced air flow rate requires a reduced fuel flow rate. The efficiency of the system is largely a function of the maximum temperature (assuming other design factors such as turbine losses are fixed). Thus, material strength considerations force the operator to reduce fuel flow rate at low ambient. When the machine is always operated at the same high temperature, then the thermal efficiency is essentially a fixed value. This is still true when you add the inlet air cooling. The benefit of the inlet air cooling comes primarily from capacity increases and not from efficiency increases.

COMBINED HEAT AND POWER

Cogeneration is a term that implies the simultaneous generation of heat and power from a fuel source. Many examples exist in practice,[18,19] including the power system here at the University of Maryland where the University recently updated its antiquated and inefficient steam generation system with a combined system that generates sufficient electric power for the entire campus plus the steam for building heat with the same amount of fuel as was previously used. The system is based on combustion turbines with heat recovery steam generators. At present, the University does not use absorption chillers although their use would be a natural way to leverage the existing steam distribution system in the summer months (the University did use absorption chillers 20 years ago in an era when the technology required considerable user expertise to avoid maintenance problems but has since replaced all of the those chillers with electric-drive machines). Many such installations exist around the world where various levels of energy integration are used at campus or municipal scales.

Recently, the term CHP (for combined heat and power) has come into use to describe energy integration at the individual building scale. In the CHP scenario, several energy technologies would be cascaded (possibly including electricity generation, heating, cooling, dehumidification, and/or hot water). Absorption chillers have a natural bottoming cycle role in such integration schemes. It remains to be seen if the CHP trend proves economical since integration implies more complexity and, hence, more maintenance and operator attention. However, the concept of energy integration existed long before the term CHP was introduced. Additionally, the potential benefits in reduced fuel usage and reduced pollutant production are expected to become more important as current resource and environmental conservation trends continue. The concept of cogeneration, implying the integration of energy systems, is well known to energy engineers. Many cities and large campuses around the world have various levels of energy integration. On a municipal scale, it is easier to solve the integration issues and to negotiate reasonable terms with the power utility. A similar idea has caught on recently to provide energy integration at the building level (under the name CHP, for combined heat and power). Absorption chillers often play a key role in energy integration schemes because they represent one of the few heat-driven refrigeration technologies. Such integration schemes have a great potential for improving the overall energy utilization by utilizing the full temperature range of a combustion process. Better utilization implies longer-lasting fuel supplies and a lower rate of environmental degradation. Thus, energy integration is expected to become more important as the economics of energy use responds to the twin drivers of global warming and fossil fuel depletion.

SOLAR AIR CONDITIONING

Absorption chillers are a natural match to solar thermal energy and solar absorption cooling has a long history of research and demonstration projects. Solar energy suffers from the same economic problem as waste heat; that is, current fuel costs make both options un-economic. Solar thermal energy collection is a demonstrated technology but it requires large collector arrays due to the low intensity level of the solar energy. Furthermore, the temperature level required by standard absorption machines is at a level where concentrating collectors are needed, and these collectors are expensive to build and maintain. All of these problems have been dealt with and the technology has been demonstrated at numerous sites, and on various scales. At the end of each project, the result was that the economic constraints hold the technology back and not the technical details.

Solar-driven absorption air conditioning is a proven technology, but until the economic environment changes significantly, it is not expected to gain any market share.

CONCLUSIONS

Absorption technology has a natural supporting role in energy integration as a bottoming cycle. As one of the most widely demonstrated, heat-fired energy technologies, waste heat-driven absorption cooling is expected to grow in importance as the value of waste heat grows in relation to other resources. Particularly, when the environmental costs associated with energy are properly charged, waste heat takes on a higher value than current practice would imply. It remains to be seen how the environmental penalties will be assessed to energy users, but increases in energy costs seem likely. When the cost of fuel use rises, then the economic incentive to utilize waste heat rises as well. Although this scenario implies many challenges for us as energy users, it also implies a brightening future for waste heat-driven absorption technology.

REFERENCES

1. Bogart, M. *Ammonia Absorption Refrigeration in Industrial Processes*; Gulf Publishing Co: Houston, TX, 1981.
2. ASHRAE Chapter 1, Thermodynamics and Refrigeration Cycles. *ASHRAE Fundamentals*; ASHRAE:Atlanta, 2005.
3. Herold, K.E.; Radermacher, R.; Klein, S.A. *Absorption Chillers and Heat Pumps*; CRC: Boca Raton, FL, 1996.
4. Herold, K.E.; de los Reyes, E.; Harriman, L.; Punwani, D.V.; Ryan, W.A. *Natural Gas-Fired Cooling Technologies and Economics*; Gas Technology Institute: Chicago, IL, 2005.
5. Alefeld, G.; Radermacher, R. *Heat Conversion Systems*; CRC: Boca Raton, FL, 1994.
6. Ziegler, F.; Kahn, R.; Summerer, F.; Alefeld, G. Multieffect absorption chillers. Int. J. Refrig.-Rev. Int. Froid **1993**, *16* (5), 301–311.
7. Ziegler, F.; Alefeld, G. Coefficient of performance of multistage absorption cycles. Int. J. Refrig.-Rev. Int. Froid **1987**, *10* (5), 285–295.
8. Bosnjakovic, F., translated by Blackshear, P.L. Jr, *Technical Thermodynamics*, Holt, Rinehart, and Winston: New York, 1965.
9. Ibrahim, O.M.; Klein, S.A. Thermodynamic properties of ammonia–water mixtures. ASHRAE Trans. **1993**, *99* (Part 1), 1495–1502.
10. Yuan, Z.; Herold, K.E. Specific heat measurements on aqueous lithium bromide. Hvac&R Res. **2005**, *11* (3), 361–375.
11. Yuan, Z.; Herold, K.E. Thermodynamic properties of aqueous lithium bromide using a multiproperty free energy correlation. Hvac&R Res. **2005**, *11* (3), 377–393.
12. Zhou, X.; Yuan, Z.; Herold, K.E. Phase distribution of the surfactant, 2-ethyl-hexanol in aqueous lithium bromide. Hvac&R Res. **2002**, *8* (4), 371–381.
13. Yuan, Z.; Herold, K.E. Surface tension of pure water and aqueous lithium bromide with 2-ethyl-hexanol. Appl. Thermal Eng. **2001**, *21* (8), 881–897.
14. Kulankara, S.; Herold, K.E. Theory of heat/mass transfer additives in absorption chillers. Hvac&R Res. **2000**, *6* (4), 369–380.
15. TICA, *Turbine Inlet Air Cooling*, 2006, http://www.turbineinletcooling.org/index.html
16. Erickson, D.C.; Anand, G.; Kyung, I. Heat-activated dual-function absorption cycle. ASHRAE Trans. **2004**, *110* (Part 1), 515–524.
17. Erickson, D.C. Power fogger cycle. ASHRAE Trans. **1995**, *111* (Part 2), 551–554.
18. Erickson, D.C. Waste-heat-powered icemaker for isolated fishing villages. ASHRAE Trans. **1995**, *101* (Part 1), 1185–1188.
19. Brant, B.; Brueske, S.; Erickson, D.C.; Papar, R.A. New waste-heat refrigeration cuts flaring, reduces pollution. Oil Gas J. **1998**, *96* (20), 61–65.

Water and Wastewater Plants: Energy Use

Alfred J. Lutz
AJL Resources, LLC, Philadelphia, Pennsylvania, U.S.A.

Abstract
Energy efficiency in existing and new wastewater treatment facilities (WWTF) can be improved by benchmarking energy use for different types of plants. Specific applications involve dissolved oxygen (DO) control, the use of adjustable-speed drives (ASDs), methane gas, heat recovery, and renewable energy. Typical savings estimates are shown to be in the range between 10 and 30%. Challenges in implementation and operation are highlighted, along with possible solutions.

INTRODUCTION

Wastewater treatment is a mature industry that has adopted tried-and-true solutions to specific challenges. In the past, the majority of these systems have performed at reasonably satisfactory levels. There is significant potential for improvement, however, in economics and in energy efficiency—both in existing and new wastewater treatment facilities (WWTF). Explanation and discussion of the potential for savings is directed toward benchmarking WWTF energy use and to specific applications in dissolved oxygen (DO) control, including the use of adjustable-speed drives (ASDs). The use of methane gas, heat recovery, and renewable energy are also discussed. Challenges in implementation and operation are highlighted, along with possible solutions.

SUMMARY

A number of opportunities in WWTF extend beyond common commercial and industrial energy efficiency and energy cost control measures. These site- and industry-specific measures often account for the bulk of the energy savings in comprehensive energy audits and retrofit projects.

Most of these measures entail control of processes on specific variables, not unlike the control of heating ventilation and air conditioning (HVAC) systems on temperature. Unfamiliarity with the application of specialized equipment and challenges in application, however, can limit their recommendation.

Nevertheless, many wastewater systems can be operated more efficiently, and the opportunities to do so must be pursued for the benefit of municipalities and private concerns that operate these systems.

Opportunities may exist in several areas. Analysis at many facilities has shown that DO control and the conversion of other speed control technologies to modern ASDs can save significant amounts of the total facility energy; typical savings estimates range between 10 and 25%. Significant additional savings can be gained by utilizing methane generated in the process, where applicable.

By gathering the necessary information and identifying appropriate partners, energy service providers and municipalities themselves can concentrate efforts on implementing the largest and most cost-effective energy efficiency measures.

GOALS OF WASTEWATER TREATMENT

The desired result in wastewater facility operations is the treatment of water to control water pollution. Certain standards and parameters are contained in operating permits for a facility; these typically are regulated and enforced by various governmental agencies. There can be substantial penalties for noncompliance.

The primary goal of WWTF is to comply with permit requirements concerning effluent water quality. These concerns—and not energy use—are paramount. Changes to systems and processes can involve savings and reduced manpower requirements but also involve some risk—especially if there are no existing areas of noncompliance. There can be substantial resistance to any changes in the way a plant operates, at many different levels.

Nevertheless, the savings opportunities are too great to wait for a major plant upgrade or total redesign to implement cost-effective energy measures, especially at older plants.

WASTEWATER TREATMENT FACILITY OPERATIONS

There are several types of treatment plants in operation, with different types of equipment. The decision to use

Keywords: Wastewater treatment; Energy use; Benchmarking; Efficiency; Dissolved oxygen; Adjustable-speed drives.

various types of equipment and plant processes is based upon several factors, including the characteristics of waste treated, variability of flows, levels of treatment needed, facility location, access to land, and proximity to the community served. A discussion of all types of plants and operations is beyond the scope of this article. The basic principles of wastewater treatment should be understood, however, to appreciate the variables and the effects of changes that may be considered.

Setting boundaries on the process, for this discussion consider first the inputs—raw wastewater, oxygen, and chemicals—and the outputs—treated effluent water and dewatered solids (sludge).

Several processes and functions cause the transformation from the input to the output stage. Although flow by gravity is desirable, most WWTF rely on a substantial number of pumps to move the wastewater to, through, and from the plant. Influent wetwells may be required. A typical first process is primary screening and settling operations. This can be followed by an aeration process, where oxygen is mechanically introduced into the wastewater or diffused into the wastewater. The oxygen is needed by microorganisms in the biological processes of transforming the incoming wastewater.

Chemicals are added to the water for various purposes. Important processes use chemicals to "coagulate" and "flocculate" solids, binding them together so that they fall to the bottom of the tank and form a sludge blanket. Then this sludge is pumped to holding, storage, or processing operations.

Some of this sludge is sent back to the aeration process. The microorganisms and bacteria responsible for the processing of wastewater possess biological activity that is based on the age of the organism, so the microorganisms must be refreshed continually. Return activated sludge is "recycled" to keep the process fresh and alive. Varying recycle rates are recommended for different types of plants, and each plant fine-tunes this measure, based on incoming flow and transit time through the plant. Some form of variable flow technology usually is employed to allow optimal recycle rates.

This return of activated sludge is a critical part of the process. The process can be "upset" and can become unstable; the process can be unable to produce good-quality effluent if too much or too little sludge is returned. A large rain event can create excess flow through the plant (due to infiltration and inflow into sewer lines or to the presence of combined wastewater/stormwater systems). In this case, the active sludge may be washed out of the plant.

Waste sludge that is not returned to the process is handled in various ways. It may be thickened and dewatered using belt presses, vacuum filtration, and centrifuges. In some cases, sludge drying beds are employed. In larger plants, sludge may also be incinerated; in this case, electricity or heat for other plant processes may be generated. The sludge also may be sent to closed digesters for aerobic and anaerobic digestion. Methane will be generated in this process. This methane can be used to produce heating for the digester operations, space heating, hot water, and electricity. In many cases, however, this methane is flared on site due to technical, maintenance, or other operational issues.

The primary output of the plant is treated water, which can be chlorinated or disinfected as needed prior to discharge. Secondary outputs (ash from sludge incineration and processed dewatered sludge solids) typically are landfilled or land-applied.

Some important terms that define the qualities and the rate of flow of influent raw water and treated effluent water are:

- DO
- Biochemical oxygen demand (BOD)
- Total suspended solids (TSS)
- Millions of gallons per day (MGD or mgd)

It is normal to measure chlorine residuals in the effluent if chlorine is used in the final disinfection process. Other measurements may entail ammonia and phosphorous.

TYPES OF TREATMENT SYSTEMS AND OPPORTUNITIES

Many facilities use a form of the activated sludge process, which, as described above, returns live microorganisms to the aeration treatment tank.

The aeration treatment tank can be aerated in several ways:

- Mechanical aeration—Large motors drive impellers to agitate water in the basin, forcing air contact with wastewater.
- Course bubble diffusion—Blowers pressurize air in headers along the bottom of the aeration basin; headers have large diffuser openings; bubbles typically are 20 mm in diameter; and agitation can be quite violent.
- Fine bubble diffusion—Blowers pressurize air in headers along the bottom of the aeration basin; headers possess ceramic heads or membrane diffusers; bubbles typically are 2 mm in diameter; contact areas are increased; and agitation is minimal.
- Pure oxygen systems—Liquid oxygen is converted to pressurized gas and fed into covered aeration tanks; motors are used for mixing and distributing oxygen throughout the wastewater.

Extended aeration refers to a system in which the aeration times are longer due to system design. This method may be preferable when there are low organic loadings (low BOD) and when space is available.

Trickling filters may also be used; basically beds of a solid, rock, or plastic medium, they provide acceptable contact time for oxygen uptake and treatment, at a low energy cost.

Advanced systems encompass nitrification, denitrification, phosphorous removal, and filtration methods, among other processes. Many municipalities currently have permit discharge requirements that dictate some form of advanced treatment.

This article concentrates on facilities utilizing secondary treatment only. Many of the results contained herein are relevant and can be applied to facilities with advanced design.

FACILITY ENERGY USE

Different types of plants will require different amounts of energy for a measure of treatment. One metric for energy use—kWh/mg—has been used to describe the energy intensity of WWTF. This metric—the kWh needed to treat each million gallons (mg) of wastewater—may be useful when correlated to a size range (e.g., 10–50 mgd) and type of treatment (e.g., secondary treatment using mechanical aeration). It may also be useful to correlate the kWh/mg with similar input and output parameters in the water treated.

From field studies that concentrated on smaller plants (primarily with average flows between 1 and 20 mgd) in the northeastern United States, the overall average for plant energy use was determined to be 1356 kWh/mg (as depicted in Table 1).

Table 1 Energy use in specific wastewater plants

Plant type no.	Average millions of gallons per day (MGD)	Peak/ design MGD	kWh/mg
1	12.0	17.0	891
1	5.0	7.1	1062
1	1.0	1.4	2174
1	5.8	10.0	1596
2	17.3	20.0	1860
2	5.8	7.0	1542
3	1.5	2.75	1193
3	8.9	10.75	1317
3	62.0	80.0	943
4	12.5	16.0	980
Average	13.2	17.2	1356

Plant type legend: 1, surface aeration; 2, course bubble diffusion; 3, finebubble diffusion; 4, high purity oxygen.

It is interesting to note the range of usage values by the types of plant. Energy use in activated sludge plants with surface aerators appears to be somewhat higher than energy use in plants with fine bubble aeration. Course bubble diffusion systems, however, appear to yield the highest energy consumption per unit of wastewater treated.

The last two columns in Table 2 show other metrics of energy use; KW/mg represents plant peak kW per average mg, whereas kW/Design mg represents maximum kW for design flows. These metrics are instructive in considering energy demand as well as consumption, and in making observations on how a plant operates relative to its loading (i.e., how variable energy use is in comparison to variable flows and how energy intensity tracks plant design flows). From these values, it appears that surface aeration may result in the highest kW energy demand for all types of plants studied.

Note, however, that the results in Table 2 are taken from a small sample size. The figures obtained should be taken as a guide and used as a frame of reference only. Further work needs to be done to compile a robust data set. This article presents a methodology and approach that can be used to create this body of data.

The use of kWh/mg and kW/mg for wastewater treatment is a valuable metric—an energy use index that can be used to set targets for energy use and identify the plants that have the greatest opportunities for savings. The results also could be used with a portfolio of plants, similar to the U.S. Environmental Protection Agency's ENERGY STAR program. One of the goals of that program is to benchmark facilities across a portfolio of commercial or institutional properties operated or managed by a single entity, to identify the best candidates for energy efficiency upgrades.

Many studies contain backup information regarding the value of kWh/mg. An Electric Power Research Institute (EPRI) study[1] cites a figure of 1212 kWh/mg for plants using secondary treatment. The same study cites figures of:

- 955 kWh/mg for plants with trickling filters
- 1322 kWh/mg for plants using activated sludge (no breakdown for type of aeration process)
- 1541 kWh/mg for plants with some advanced treatment (but without nitrification)
- 1911 kWh/mg for plants with advanced treatment and nitrification

It is noted that some facilities operate at significantly higher levels of energy intensity; some advanced or extended aeration plants with nitrification have recorded intensities that exceed 3000 kWh/mg. Results indicated in Tables 1 and 2 pertain primarily to facilities that utilize only secondary treatment processes.

Size of facilities can be important when using kWh/mg as a measure of energy intensity or retrofit potential.

Table 2 Energy use ranges by type of aeration process

Plant type	Value type	kWh/mg	kW/mg	kW/design mg
All Types	Average	1356	94.0	67.9
	Maximum	2174	182.0	130.0
	Minimum	891	51.4	40.2
Surface	Average	1431	119.8	80.4
Aeration	Maximum	2174	182.0	130.0
	Minimum	891	71.8	50.7
Coarse	Average	1701	97.6	83.0
Bubble	Maximum	1860	116.5	100.8
	Minimum	1542	78.6	65.1
Fine	Average	1151	71.6	50.4
Bubble	Maximum	1317	83.3	64.1
	Minimum	943	53.9	41.8

Studies conducted by the United Nations[2] and others indicate that results vary for different sizes of plants. In ranges of 1–10 mgd, for example, results from Germany indicated energy use for activated sludge plants of approximately 1600 kWh/mg. For plants smaller than 1 mgd, consumption was approximately 1900 kWh/mg, whereas for plants larger than 10 mgd, consumption was approximately 1200 kWh/mg. This large variation illustrates the importance of considering plant size when using kWh/mg as a comparator.

Interestingly, unit consumption was reported as sometimes increasing with plant size, perhaps due to higher standards of treatment. Other countries in this study, including Canada and the United Kingdom, showed very low energy consumption figures in wastewater treatment. Again, this may be for different levels of treatment required or the use of other treatment technologies.

Electric Power Research Institute also reports a sizable increase in energy intensity in facilities that are below the 5 mgd range, as compared with larger facilities.[1]

Amounts of energy use by process for a typical activated sludge secondary treatment plant are estimated[1] as follows:

- Aeration: 47%
- Sludge processing: 32%
- Pumping: 16%
- Balance of plant: 5%

These estimates of energy use by process are only approximate figures. One particular case study of a 34.1 mgd plant has shown energy use in aeration systems as high as 73% of total plant energy.[4] Field investigation and measurement are required to determine reasonable and valid estimates of energy use at individual facilities.

These percentages can be used, however, to indicate the processes in which it is possible to find large savings opportunities—namely, aeration, sludge processing, and pumping.

ENERGY EFFICIENCY MEASURES

Large opportunities exist for efficiency in WWTF (see Fig. 1). Many of the larger systems will interact with smaller systems, and opportunities can be implemented concurrently. A detailed analysis should entail all processes and systems within the facility.

Dissolved Oxygen Retrofits and Control

Dissolved oxygen is required due to the BOD of the incoming untreated wastewater. Dissolved oxygen is introduced into the wastewater by several means. The most common techniques are through the use of mechanical surface aeration or aeration blowers.

Table 3 shows estimated savings at several WWTF studied. The savings are attributed to different measures affecting aeration systems, given as a percentage of total facility kWh consumption. The percentage of kWh savings

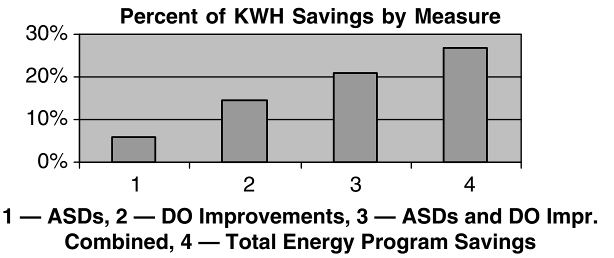


Fig. 1 Example of savings due to energy efficiency.

Was–Wat

Table 3 Energy savings from aeration system modifications

Plant type no.	Pre-retrofit KWH/yr	Aeration system savings (%)	Savings from all measures (%)	Notes
1	3,901,200	7	31	Dual speed motors
1	1,938,300	4	34	—
1	793,600	6	20	Weir control
1	3,377,900	0	25	—
2	11,747,400	18	26	Fine bubble and downsizing
2	3,265,000	14	20	Fine bubble and downsizing
3	652,900	23	23	Dissolved oxygen (DO) Control and adjustable-speed drives (ASDs)
3	4,278,960	0	6	Process control savings
3	21,340,000	15	27	Downsizing and guide vane control
4	4,471,400	21	27	ASDs on aeration motors
Average	5,576,666	11	24	—

from all energy efficiency measures identified, as a percentage of total kWh consumption, also is given to illustrate the contribution of aeration system modifications in a comprehensive energy efficiency program.

In mechanical surface aeration systems (plant type no. 1), a number of motors will be used to drive large impellers. These motors agitate the surface of the water, continually entraining oxygen. Motors may be left on continuously or energized/deenergized manually or automatically, dependent on operator experience or lab results. There are several possible retrofits to these systems.

A seemingly simple retrofit would be the conversion from single-speed to two-speed motors. There is also a possibility of using ASDs. Some considerations include oil circulation in gear boxes at reduced speed; tank geometry and the possibility/consequences of short-circuiting of wastewater flows, impacting treatment; location of motor control centers and electronic variable-speed drive placement to prevent damage due to severe ambient conditions and to minimize impact on facility power quality; and the ability of existing motors and their insulation systems to handle reduced speeds.

It is also possible to reduce energy use of surface mechanical aerators by changing the submergence of the driven impellers through weir control or other means. In this case, motor loading will decrease. Again, short-circuiting may need to be evaluated. In general, automatic control should yield the greatest savings, although a manual weir control system could be implemented. This strategy has been reported in a case study to yield savings of 12%–15% in total plant energy.[4]

With automatic control systems, sensors that measure control parameters are needed. Dissolved oxygen is the most common control parameter. A number of DO sensors are available. Newer technologies include self-cleaning sensors. These can be important; sensors lose accuracy as they are subject to fouling in unclean environments. Sensor location is important; there typically are many sensors per tank. Routine calibration also is essential to process control and achieving energy savings.

The inputs from DO sensors and the output signals to weirs, motors, or ASDs can be fed into the plant's computerized control system.

The Environmental Protection Agency (EPA) study cites that the use of better DO control can save significant amounts in older facilities, as shown in three case studies: 12.5% in a 90 mgd plant, 14.5% in a 53 mgd plant, and 57% in a 36 mgd plant.[3] These results were measured in pounds of BOD removed per kWh used, which ranged from a low of 0.88 lb/kWh to a high of 5.95 lb/kWh. Other studies concur that savings from automatic control of 8% to more than 20% are possible.

A common aeration technique involves the use of blowers and pressurized air distributed through tanks via headers. The blowers may be the centrifugal or the positive-displacement type. Normally, valves or guide vanes would be used to regulate airflow to the manifolds and headers. The valves or vanes generally would be controlled on inputs from DO sensors. In some cases, there is the ability to optimize these controls.

In other cases, instead of throttling flow via valves, which introduce friction and lead to energy losses, ASDs can be used. Positive-displacement blowers and centrifugal blowers lend themselves to the use of ASDs.

The application to centrifugal blowers requires that turndown ratios be regulated to prevent blower surging. Airflow must be maintained above the minimum standard cubic feet per minute (SCFM) required by the diffusers in the system.

In addition to ASDs to control blower energy use, stepped motors can be considered for new facilities, and possible motor downsizing should be considered in existing facilities. It was forecast that facilities in a California program would be able to obtain 39 and 63% energy savings by using smaller blowers in aeration systems.

Course bubble systems can be inefficient due to large bubble size, low oxygen transfer efficiency, the requirements for blowers, and violent agitation of water in the aeration basins. Many plants have been retrofitted to fine bubble diffusion systems. A range of 9%–40% savings on aeration system energy use by converting from course to fine bubble diffusion has been documented.[1]

Various systems have been used for fine bubble diffusers. Ceramic disks were some of the first available systems; membrane systems are also available. There are efficiency advantages to membrane systems; oxygen transfer efficiencies of 13 vs 10% have been cited.[5] Membrane systems may also have the ability to be cleaned by "pulsing," or rapidly changing airflow. Cleaning and maintenance are important considerations in the selection of diffuser heads for use in specific facilities.

Adjustable-Speed Drives

There are many pumping systems within a wastewater treatment facility. These include influent wetwell pumps, return-activated sludge pumps, waste-activated sludge pumps, effluent pumps, and plant water systems.

The need to regulate variable flows and speeds at many stages in the process derives from the need for stable plant operations to produce high-quality treated effluent. Because pumping also consumes a large quantity of energy, better control of all pumping processes should be considered. In many cases, a reduced cycling schedule or better pump impeller selection for changed conditions may be relatively simple, low-cost improvements. The use of variable-speed pumping and ASDs to replace older flow control technology, although having possible higher capital costs, has proved to be advantageous both in reducing energy use and in improving overall process control.

It is possible to replace certain existing speed control technologies with electronic or mechanical ASDs. With addition of control system components, process control can be optimized.

Eddy current drives, for example, have been used to reduce speeds and flows mechanically through return sludge pumps in many plants. Flow generally is set at a certain recycle percentage. During periods of high or low flow, that percentage may require manual adjustment. By linking a process controller with an influent flow meter, an ASD will be able to select a variable optimized setting automatically—a process known as flow pacing.

Similarly, larger pumps may have been controlled with liquid rheostat-type controls. These controls can regulate flow but do not match the efficiencies of electronic ASD.

Table 4 shows estimated savings due to application of modern electronic variable-speed technologies, in kWh and as a percentage of total facility kWh consumption. As in Table 3, the percentage of kWh savings from all measures as a percentage of total kWh consumption also is given to illustrate the contribution of speed control modifications to a comprehensive energy efficiency program.

Care must be taken to locate these drives with the correct enclosures, provide a suitable operating environment, and check compatibility with driven motors and pumps. Analysis of the pump curves is essential.

Energy savings calculations must take into consideration the efficiencies of existing systems at various flow ranges. In many cases, the existing motor systems draw less than the motor full load horsepower. This can lengthen the payback on energy-saving retrofits. Detailed study

Table 4 Energy savings from speed control modifications

Plant type no.	Pre-retrofit consumption kWh/yr	Speed control savings	kWh Savings — % of total (%)	Total savings of all measures (%)
1	3,901,200	607,300	16	31
1	1,938,300	450,669	23	34
1	793,600	64,800	8	20
1	3,377,900	409,629	12	25
2	11,747,400	107,900	1	26
2	3,265,000	119,961	4	20
3	652,900	—	0	23
3	4,278,960	—	0	6
3	21,340,000	1,304,058	6	27
4	4,471,400	—	0	27
Overall	5,576,666	437,760	8	24

Was–Wat

supplemented by field measurements should be used to verify the efficacy of these retrofits.

Sludge Processing and Methane Gas Recovery

Sludge processing consumes a significant amount of energy. Dewatering and digestion of the sludge can be an energy-intensive process. One must consider not only the energy used in processing sludge, but also the energy value of the byproducts of this process. Sludge can be incinerated, and methane gas produced in digestion can be used. Both byproducts can generate electricity and heat, which can be used in plant operations. Conversion from vacuum filters to more efficient technologies should be evaluated. Solar sludge drying, for example, also may prove to be cost effective.

Methane is formed in the anaerobic digestion of sludge. In some cases this valuable fuel is wasted, being flared to the atmosphere. Some facilities flare the total quantity of gas generated, whereas others use methane as a source of process, space, and water heating through on-site boilers.

In many cases it is economical to capture this methane; treat the gas by drying and removing impurities; and use this gas in an internal-combustion engine, in gas turbines, or even as feedstock for fuel cells.

In one instance, in a project funded and financed by an energy service company, a New Jersey municipal authority saved more than $450,000 (estimated as approximately 20% of its aeration costs) by using a gas-fired engine to drive a new blower directly.[6]

In another case, a smaller plant with flows of less than 2 mgd expects to save more than 200,000 kWh/yr, or approximately 30% of total plant electrical energy use. Not only will the system provide some backup power, but also, the savings for this small municipality will be more than $10,000 per year.

Sludge also has been incinerated to reduce volume and landfill costs. On-site incineration generally applies to larger systems. Sometimes it is cost effective to install heat-recovery boilers, which can be used to generate electricity and provide heat for both processes and winter space heating. One such project at a large municipal WWTF with a 40 mgd design flow has been estimated to save $1.2 million annually.

Renewable Energy Measures

There have been a few applications for other renewable or alternative energy sources. Solar sludge drying has potential applications where space and odor problems can be addressed.

Another approach for larger lagoons involves the use of introducing slow-rate laminar flows through motors that can be powered by solar electric (photovoltaic) cells. These systems increase the oxygen transfer to the system at a very low energy cost.

A few municipalities with high elevation drops on the outflow from WWTF (or in related water systems) have the potential to use simple hydroelectric turbines. This measure applies to few systems, as the elevation potential (in feet of head) usually is not sufficient to make this approach economical compared with current power purchase prices and electric tariffs.

Other Energy Cost Control Measures

Other process areas that should be addressed include the optimization of control systems, wetwell level adjustments to increase suction head, and chlorine system modifications. As an alternative to using variable-speed drives or dual-speed motors, some facilities may be better served by downsizing motors to accommodate low flow periods or actual operating conditions. One area where this has been implemented is in the optimization of plant water systems.

Due to the number of motors, it may make economic sense to retrofit using premium high-efficiency motors. In motor replacement, the highest-efficiency motor for a specific application should be selected, as the incremental costs of this motor can yield a short payback, sometimes paying for itself in a few months.

To reduce energy costs, hardware and software changes may be of benefit in improving power factor, limiting peak kW demand, and transferring load to off-peak rates through time-of-use scheduling.

Another area that deserves attention is the use of chemicals to reduce loading and energy requirements in secondary processes. This was implemented at a facility designed for 216 mgd that served a large North American city.[4]

Proper equipment selection and facility design enable significant energy savings. A savings of slightly more than 20% in oxygen uptake—a difference of 1.59 vs 1.26 kg of oxygen per kWh consumed—was reported at different plants within the same facility.[4] This applies to the design of new plants and in substantial expansions/renovations.

In support buildings, HVAC and lighting improvements should be considered. Due to the high hours of use in some plants, facilitywide lighting retrofits usually are cost effective.

CONCLUSIONS

There are enormous opportunities to save energy in existing WWTF and to optimize design in newer facilities. Many plants can save on the order of 10%–30% of energy costs through cost-effective retrofits. Baselines of energy usage have been developed for different types of plants. This benchmark for estimating possible savings in a particular facility may be useful. Process operations with high energy use, particularly the DO process, should be given special attention. Modern speed control technologies, such as electronic or mechanical ASDs, have relevance in reducing operational costs and in improving process control. Both energy and environmental benefits accrue when renewable energies are utilized for some of the energy requirements. In particular, methane recovery and use are particularly appropriate for those facilities with digesters and in the design of new facilities.

REFERENCES

1. Electric Power Research Institute. *Water and Wastewater Industries: Characteristics and DSM Opportunities. EPRI TR-102015*; Burton Environmental Engineering: Los Altos, CA, 1993.
2. United Nations. *Strategies, Technologies, and Economics of Wastewater Management in ECE Countries*; 1984.
3. U.S. Environmental Protection Agency. *Upgrading Existing Wastewater Treatment Plants-Case Histories. EPA-625/4-77-005a*; 1973.
4. Diagger, G.T.; Buttz, J.A. *Upgrading Wastewater Treatment Plants*; Technomic Publishing, Inc.: Lancaster, PA, 1998.
5. Iranpour, R.; Stenstrom, R.K. Relationship between oxygen transfer rate and flow for fine pore aeration under process conditions. Water Environ. Res. **1991**, *73* (3).
6. Polaski, J.; Schmalz, G. Energy-efficient retrofit saves electricity and money for wastewater plant. Water Eng. Manag. **2000**, *147* (1), 28–29.

Was–Wat

Water and Wastewater Utilities

Rudolf Marloth
Department of Mechanical Engineering, Loyola Marymount University, Los Angeles, California, U.S.A.

Abstract

In the first half of the 20th century, life expectancy in the United States increased dramatically, primarily because of water treatment, which greatly reduced the incidence of waterborne bacterial infections such as cholera, typhoid fever, and dysentery. The Safe Drinking Water Act of 1974 and its amendments of 1986 and 1996 are the primary pieces of federal legislation protecting drinking water supplied by public water systems. A standard process chain for water treatment consists of grit removal, flash mixing with chemicals, flocculation, sedimentation, filtering, and disinfection.

The Clean Water Act of 1972 and subsequent legislation with the objective of reducing the discharge of pollutants to natural waters imposed standards on the secondary treatment of wastewater. A standard process chain for wastewater treatment consists of preliminary (physical), primary (physical and chemical), and secondary (biological) steps. In secondary treatment, microorganisms consume organic pollutants. Much industrial waste is incompatible with public treatment systems, so it is usually subject to pretreatment.

INTRODUCTION

Municipalities are normally responsible for providing water and disposing of wastewater, either by privately or publicly owned utilities. A water utility is responsible for the source, transmission, treatment, and distribution of potable water. Water usage is metered at the user. A wastewater treatment utility is responsible for collection, treatment, and disposal. Industrial wastewater is often metered. Residential wastewater charges are based on water usage.

Municipalities are also responsible for the collection and disposal of runoff, which consists of stormwater, misplaced or excessive irrigation, domestic car washing, etc. Runoff collects in gutters or holding ponds; then goes to storm sewers; and is dumped into a nearby river or, in coastal cities, directly into the ocean. This water carries with it a variety of pollutants: fertilizer, animal waste, oils, tire dust, etc. For municipalities that do not have combined storm and sanitary sewers, this runoff is dumped without treatment. Worse, in severe storms, combined sewer systems are likely to dump runoff combined with untreated sewage. Furthermore, runoff is not metered, so the municipality bears the cost of installing and maintaining the infrastructure. Possible approaches to reducing pollution of natural waters are (a) enactment of ordinances to make property owners responsible for reducing runoff and its pollutants, (b) diverting light flow or the first flow in a storm to a wastewater treatment plant (WWTP), and (c) installing treatment facilities to make runoff suitable for reuse. (The city of Santa Monica, California recently demonstrated its SMURRF (Santa Monica Urban Runoff Recycling Facility), the first of its kind in the nation.)

Keywords: Activated sludge; Baffles; Broad street pump; Clean Water Act; Coagulation; Coliforms; *Cryptosporidium*; Disinfection; Filter cycle time; Filtering; Flash mixing; Flocculation; *Giardia*; Head; Hydraulic; Industrial waste; Log removals; Microorganism; Natural waters; Pollutants; Preliminary; Primary; Reactor; Safe Drinking Water Act; Secondary; Sedimentation; Settling velocity; Sewage; Slow sand filter; Sludge; Tertiary; Turbidity; Wastewater; Water.

WATER

The primary objective of water treatment is to make it safe for human consumption at a reasonable cost. It is possible to produce safe water that has objectionable taste, odor, or color, so a secondary goal is to make the water appealing to the consumer. Turbidity and color are qualities apparent to the naked eye. The former is caused by particles in suspension and is measured in nephalometric turbidity units (NTU) by passing light through a sample. These particles can be removed by settling or filtering. Colloidal particles will not settle in a reasonable amount of time and must be removed by other physical processes. Dissolved substances are removed or transformed by chemical treatment. Color may be caused either by materials in solution (true color) or in suspension (apparent color).[1]

History

Early water treatment focused on what was apparent to the senses: appearance, taste, and odor. These qualities were improved by removing turbidity through filtration or precipitation. It was found much later that particles in the water harbored pathogens, which were largely removed while clarifying the water.

A slow sand filter (SSF) designed by James Simpson was commissioned in 1829, but it was some time before its full importance was realized. This simple device is essentially a tank filled with sand with water introduced at the top and removed at the bottom. Some time after the filter is started up (a few days to a few weeks), the upper layer of sand becomes coated with a gelatinous biological layer called the schmutzedecke, made up of algae, bacteria, protozoa, and small invertebrates. This sticky layer is biologically active and converts organic matter in the water to water, carbon dioxide, and harmless salts (i.e., it is mineralized). Later, research showed the importance of biological removal and also showed that the SSF was very effective in that respect. In particular, the SSF is effective in removing *Giardia* and *cryptosporidium* oocysts, which are nearly unaffected by chlorination.[1,2]

In the famous Broad Street pump episode of 1854, an outbreak of cholera in the Soho district of London killed more than 600 people. Dr. John Snow, who had theorized that the disease was spread by contaminated water, traced it to water from the Broad Street pump. The likely cause was from a leaking and cholera-infested cesspool located only three feet from the Broad Street well. In fact, cesspits lay under many of the houses in the district.[3] Acceptance of Snow's theory was slow, as was conversion to a sewer system that conveyed wastewater to central plants.

Recently, the Centers for Disease Control and Prevention and the National Academy of Engineering named water treatment as one of the most significant public health advancements of the 20th century.[4] This is rightly so, for in the first half of the 20th century, life expectancy in the United States increased dramatically, primarily because of water treatment, which has greatly reduced the incidence of waterborne bacterial infections such as cholera, typhoid fever, and dysentery. Even so, waterborne disease does occur in this country, the most well-known instance of which is the cryptosporidiosis outbreak of 1973 in Milwaukee. Most episodes are due to contamination of raw or treated water, inadequate treatment, and cross-contamination between sewers and water mains.

Standards and Monitoring

The Safe Drinking Water Act of 1974 and its amendments of 1986 and 1996 are the primary pieces of federal legislation protecting drinking water supplied by public water systems. Primary regulations under the act are for the protection of public health; secondary regulations are for regulations pertaining to taste, odor, and appearance.

The Surface Water Treatment Rule mandates that surface water or groundwater under the influence of surface water must be treated to remove or inactivate 99.9% of *Giardia lamblia* cysts and 99.99% of enteric viruses. These requirements are commonly stated as "log removals," where n-log removal is removal of a $1–10^{-n}$ fraction of the pollutant. Treatment processes using filtration are judged to comply by providing an adequate concentration/contact time (Ct) product, where C is the concentration of the disinfectant in mg/L and t is the time in minutes.[5]

The Enhanced Surface Water Treatment Rule requires a 2-log removal of *Cryptosporidium*. Systems using filtration are granted credit if they meet certain turbidity criteria. The Filter Backwash Recycling Rule requires treatment plants to recycle filter backwash water through the entire process cycle. The Disinfectants/Disinfection Byproduct Rule establishes maximum contaminant levels (MCLs) on total trihalomethanes, haloacetic acids, bromate ion, and chlorite ion.[5]

The most common way of testing the quality of drinking water is the coliform test, as specified by the U.S. Environmental Protection Agency. Coliforms are bacteria that are gram-negative, aerobic and facultative anaerobic, nonspore-forming rods, which ferment lactose with gas formation in 48 h at 35°C. When a sample tests positive for coliforms, it must be tested for fecal coliforms. Fecal or thermotolerant coliforms include all coliforms that can ferment lactose at 44.5°C. It is common to identify *Escherichia coli* uniquely with fecal coliforms.[5]

Sources and Transmission

Large-scale sources are primarily surface water and groundwater. Desalinated ocean water is not yet a major source in the United States, although Spain is a major producer and user of desalinated water. In the past 30 years, the energy required for desalination has fallen from 12 to 3 or 4 kWh/m^3 using reverse osmosis.[6] In a few areas, recycled water (treated wastewater) is used for irrigation. Water is transmitted by way of natural water courses, lined open channels, or pressure pipe. Los Angeles, for example, receives water from the Sacramento Delta via a concrete-lined open channel, from snowmelt in the Sierra via a natural watercourse that is diverted into an iron pipe aqueduct. New York reservoirs supply water to treatment plants through a series of underground tunnels. In some cases, water treatment plants are located right at the site where water is drawn from a river or lake.

Treatment

Treatment is tailored somewhat to the characteristics of the influent and the effluent limits, but in general, large particles are removed first by screening and then by settling. Next are particles that will float or settle with some assistance, such as air floatation or mixing with coagulants. Colloids and dissolved materials are removed last. The water should be nearly clear before disinfection so that contaminants may not hide in turbidity.

Standard water treatment (Fig. 1) consists of coagulation, flocculation (aggregation into a wooly mass), sedimentation, filtration, and disinfection. The size of the

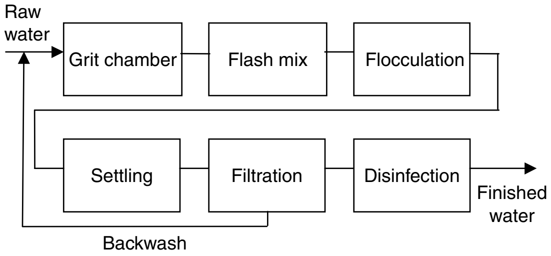


Fig. 1 Water treatment.

units and major equipment are determined by the hydraulic loading. The examples below are based on a plant with an average flow of 24 mgd (million gal per day). Plants are commonly designed for a maximum daily flow that is 1.5 times the average daily flow. Plants whose main supply is groundwater will have a somewhat different process chain because the water will contain less settleable and suspended solids but probably more dissolved metals. Plants that draw from a river should have at their head a coarse screen for tree branches, etc., and a grit chamber for sand and silt, because this kind of grit will cause a serious maintenance problem with pumps.

The coagulation operation consists of the addition and flash mixing of chemicals meant to remove the charge on colloids and suspended solids so that they will aggregate in the flocculation step. Depending on the type of chemical used, flash mixing should occur in about one to five seconds. The addition of more than the optimum amount of chemicals can compensate for less-than-optimum mixing. A potential problem in this step is the clogging of feed lines. Various methods of flash mixing are available, the best of which are diffusion mixing by water jets and inline static mixing. Proper operation at this step requires choosing the right chemicals and quantities in response to the raw water quality and flow rate. Chemicals used in this step include aluminum sulfate (alum), polyaluminum chloride, various iron compounds, polymers, and bentonite.[5]

Flocculation is slow mixing that increases the rate of collisions between particles whose electrostatic repulsive forces have been neutralized in the coagulation step. Now the particles will stick together into sizes that will settle. Flocculation mixing is performed by mechanical mixers, baffles, or other methods. Paddle reels with a horizontal axle perform well. In a reasonable design for two parallel, 12-mgd floc tanks, each of three chambers in a tank might be about 13.5 ft^2 in cross section so that a 13-ft horizontal paddle reel would nearly fill a chamber. Perforated baffle walls separate the chambers to promote good mixing. Proper operation of this step requires continual monitoring of and adjusting for the floc size, and removing scum from the surface of the water, sludge from the bottom of the tank, and algae from the vertical surfaces (walls and baffles). Transfer of the flocculated water to the sedimentation tank must be done at low velocity to avoid breaking up the floc.

Four progressive stages of sedimentation are distinguished. Type I sediment consists of separate, destabilized particles. Type II is made of larger groups of flocculated particles. In Type III, the particles have formed a blanket that initiates hindered settling. Type IV settling is compression of the sludge blanket at the bottom of the tank.

A typical filter bed is made up of a layer of sand and a layer of coal, charcoal, or granular activated carbon. The water from the sedimentation tank is introduced at the top of the filter and moves by gravity down through the media to the underdrain. It is most desirable to have the filter backwashed once per day. Thus, the design depends upon the quality of the raw water, the required throughput, the local climate, and the skill level of the operators. The backwash water is required to be recirculated to the head of the plant.

Filter efficiency is determined by the unit filter run volume (UFRV), which is the ratio of the amount of water processed during a filter cycle to the amount that could be processed if no backwashing were necessary. The effective filtration rate, R_e, is

$$R_e = (\text{UFRV} - \text{UBWV})/T$$

where UBWV (unit backwash volume), gal/ft^2; T, filter cycle time, min; and

$$\text{UFRV} = V_f/A \qquad \text{UBWV} = V_b/A$$

where V_f, volume filtered per filter cycle, gal; A, area of filter, ft^2; V_f, volume of backwash water, gal.

The design filtration rate, R_d, is the maximum filtration rate, which can be achieved only if no backwash were necessary. Then the production efficiency is

$$R_e/R_d = (\text{UFRV} - \text{UBWV})/\text{UFRV}$$

Example 1. Find the production efficiency and the filter cycle time for a filter with UFRV equal to 7500 gal/ft^2, UBWV equal to 200 gal/ft^2, and design filtration rate equal to 5 gpm/ft^2.

Solution. The effective filtration rate, filter efficiency, and filter cycle time are

$$R_e = (5\ \text{gpm/ft}^2)[(7500\ \text{gal/ft}^2) - (200\ \text{gal/ft}^2)]\big/(7500\ \text{gal/ft}^2) = 4.87\ \text{gpm/ft}^2$$

$$R_e/R_d = [(7500\ \text{gal/ft}^2) - (200\ \text{gal/ft}^2)]/(7500\ \text{gal/ft}^2) = 0.973$$

and

$$T = [(7500\ \text{gal/ft}^2) - (200\ \text{gal/ft}^2)]/(4.87\ \text{gpm/ft}^2) = 1499\ \text{min}$$

Note that the filter cycle time is very close to one day (1440 min). Also, increasing the filtration rate will not necessarily increase the amount of water filtered per day.

Increasing the rate increases the amount of deposition on the filter media, reducing the filter run time.

Most of the water treatment plants in this country disinfect with chlorine. Common alternatives are ozone, chloramine, and ultraviolet light. The disinfectant is sometimes added at the head of the plant to give an adequate concentration-contact time product. When chlorine is used, disinfection byproducts can be formed—most notably, trihalomethanes. To suppress this, ammonia might be added at the end of sedimentation to form chloramines. Then the water is lightly rechlorinated in the clearwell to suppress regrowth of pollutants in the distribution system.

In small municipalities across the heartland, treated water is pumped into elevated tanks emblazoned with the name of the town. These towers serve to keep the water clean, meet surges in demand, and supply even pressure at the tap. Large cities tend to keep the water in open reservoirs, an unfortunate but perhaps necessary practice that leaves the water subject to the reintroduction of various undesirable substances.

Hydraulics

Flow through the plant is described by Bernoulli's equation, where all the components are expressed in terms of head in feet,

$$z_1 + P_1/\gamma + V_1^2/2g = z_2 + P_2/\gamma + V_2^2/2g + H_L$$

and z_i, distance of water level above datum; P_i/γ, pressure head at surface; $V_i^2/2g$, velocity head; H_L, head loss; $i=1$, upstream; $i=2$, downstream.

The head-loss term is made up of entrance and exit losses, pipe friction losses, and minor losses. Hydraulic calculations are best made starting at the clearwell and working upstream.[1]

It is preferable to have the water flow through the plant by gravity. The head losses through the processes are approximately (a) rapid mixing, 1 ft; (b) flocculation, 2 ft; (c) sedimentation, 2 ft; (d) and filtration, 10 ft, for a total of 15 ft. If the water comes into the plant at the level of the clearwell, this head and the friction losses in the lift pipe must be provided by a pump. At the pump discharge, the water horsepower is HP_w; the pump input is the motor brake horsepower, HP_b; and the power required to drive the motor is P, where

$$HP_w = QH/C_1 \qquad HP_b = QH/C_1 e_p$$

$$P = C_2 HP_b/e_m$$

and Q, pump flow rate, gal/min; H, pump discharge head, ft; C_1, constant, (550 ft-lb/s-hp) (60 s/min)/(8.34 lb/gal) $\approx$ 3960 ft-gal/min-hp; e_p, pump efficiency; C_2, constant, 0.746 kW/hp; e_m, motor and drive efficiency.

Example 2. Size the motor, and find the input power required to provide 15 ft of head for a 24-mgd water treatment plant. Ignore plumbing losses. Take the pump efficiency to be 70% and the motor/drive efficiency to be 90%.

Solution. The flow rate is $(24\times10^6$ gal/da)/[(24 h/da) (60 min/h)] = 16,667 gal/min. Thus

$$HP_b = QH/C_1 e_p = (16{,}667 \text{ gal/min})(15 \text{ ft})/[(3960 \text{ ft-gal/min-hp})(0.70)] = 90 \text{ hp}$$

and

$$P = C_2 HP_b/e_m = (0.746 \text{ kW/hp})(90 \text{ hp})/(0.90) = 75 \text{ kW}$$

Because a motor is just as efficient at 75% load as at full load, a 125-hp motor should be installed.

The important hydraulic parameters of a sedimentation tank (Fig. 2) are the hydraulic retention time (HRT), the horizontal velocity of the water through the tank (V_h), and the surface overflow rate (SOR). For a tank with dimensions L, W, D for length, width, and depth, the volume $\forall$, is LWD; the surface area, A_s, is LW; and the vertical cross-sectional area is $A_c = WD$. Therefore, the horizontal velocity and the SOR are

$$V_h = Q/A_c = Q/DW \qquad \text{and} \qquad SOR = Q/A_s$$

The HRT is

$$HRT = L / V_h = L / (Q / DW) = \forall / Q$$

A particle with settling velocity satisfying HRT $= D/V_s$ will reach the bottom of the tank before being swept up and out (as will particles with V_s greater than this). Then

$$V/Q = D/V_s \qquad \text{or}$$

$$V_s = DQ/V = Q/LW = SOR$$

That is, for a particle to settle, the settling velocity must be equal to the SOR (or greater).

The power (P) required for mixing or flocculation with an impeller in a tank is dependent upon the average

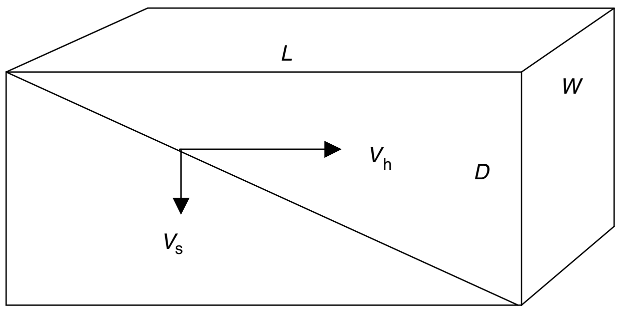


Fig. 2 Sedimentation parameters.

velocity gradient (G) in the fluid, dynamic viscosity (μ) of the fluid, and volume ($\forall$) of the tank: $P = G^2 \ \mu \ \forall$. For wastewater treatment, the average velocity gradient for rapid mixing is about 500–1500/s and for flocculation is about 50–100/s.[5]

Example 3. Find the power (P) required to achieve an average velocity gradient (G) of 100/s in a 1 million-gal flocculation tank whose contents are at a temperature (T) of 40°F.

Solution. The dynamic viscosity of water at 40°F is $\mu = 2.359 \times 10^{-5}$ lb-s/ft². Therefore,

$$\begin{aligned} P &= (100/\text{s})^2 (2.359 \times 10^{-5}\text{lb–s/ft}^2)(10^6 \text{ gal}) \\ &\quad \times (0.746 \text{ kW/hp})/[(550 \text{ ft–lb/s–hp})(7.48 \text{ gal/ft}^3)] \\ &= 42.8 \text{ kW} \end{aligned}$$

Example 4. Consider a plant with an input flow rate of $Q = 24$ mgd $= 24 \times 10^6$ gal/da. The two sedimentation tanks are 300 ft long, 40 ft wide, and 13 ft deep. Find the HRT, the velocity of the water through the tank, and the SOR.

Solution. The flow rate through each tank is $Q = (24/2 \times 10^6$ gal/da)/[(24 h/da)(60 min/h)] = 8333 gal/min = (8333 gal/min)/[(7.48 gal/ft³) (1440 min/da)] = 1114 ft³/min. Therefore, the velocity in each tank is $V =$ (1114 ft³/min)/[(40 ft) (13 ft)] = 2.14 ft/min, and the surface overflow rate is SOR = (8333 gal/min)/[(40 ft) (300 ft)] = 0.694 gal/min-ft².

WASTEWATER

Some authors make the following distinction: wastewater is water that has been used for domestic, industrial, or commercial purposes, whereas sewage is more inclusive in that it can include water that has not been used, such as rain runoff. In the past 50 years, sanitation agencies have made a great effort to confine runoff to storm sewers and out of treatment plants, so the term WWTP is appropriate. These facilities are also called publicly owned treatment works (POTWs).

History

The history of wastewater treatment is a sordid one of determined ignorance and apathy. Until 1965, for example, Salt Lake City was dumping raw sewage into a 9-mi open canal that emptied into the Great Salt Lake.[7] In other parts of the world, this kind of practice continues to the present day.

Dr. John Snow in London convincingly linked cholera with the consumption of contaminated waters. This most famous episode in the history of both epidemiology and water treatment occurred in the late summer of 1854. Repeated outbreaks of cholera had occurred between 1831 and 1854 in the industrial cities of England, with little being done to prevent or contain it. The particularly sudden and violent episode in and around Broad Street in September gave Dr. Snow the opportunity to verify his belief that the cause was in contaminated water. When he persuaded the authorities to remove the handle from the Broad Street pump, the spread of the disease was halted. Later, Snow established that wastes from a single infected individual had been dumped into a leaking cesspit near the Broad Street well. After some time, people accepted the fact that fecal contamination of drinking water was a major cause of disease.[3]

With better water supplies and sewer systems, there was a sharp decrease in the incidence of waterborne diseases, even before the agents were identified. After half a century of research, the concept of waterborne disease was established. The cause was known to be microorganisms in the digestive tract and the associated health hazards had been proven. Then work proceeded on two fronts: analytical methods for the detection of fecal pollution and the development of treatment methods and facilities. Research led to publication of the first edition of *Standard Methods* in 1901,[8] and the SSF was an early and very effective method for treatment. (In fact, the newer and faster conventional rapid sand filter does not remove dissolved constituents as effectively.)

Standards and Monitoring

The Clean Water Act of 1972 and subsequent legislation placed increased emphasis on the importance of reducing the discharge of pollutants to natural waters. The minimum national standards now for secondary treatment are the "30/30" rule: a 30-day average of no more than 30 mg/L of BOD_5 (5-day biochemical oxygen demand) and TSS (total suspended solids), as well as pH to be between 6.0 and 9.0 at all times. Unfortunately, meeting these standards does not guarantee the absence of disease-causing agents—notably, *Giardia lamblia* and *Cryptosporidium parvum*. Increased sophistication of monitoring techniques is leading to better treatment techniques and stricter standards.[9]

Sources and Collection

Sanitary sewers receive some groundwater infiltration and stormwater. Otherwise, 90% or more of the intended influent is of residential or commercial origin. Industrial users are either direct dischargers dumping into a waterway or indirect dischargers dumping into a

POTW. (Some industrial liquid waste may also be hauled off.)

Influent is collected in closed pipe, mostly by gravity flow. When pump lift is necessary, the flow is into short runs of pressure pipe, eventually returning to gravity flow lines. Because plants are often located next to natural waters, they are typically at the low points of terrain, which keeps pumping to a minimum.

Treatment

Several levels of increasing care are defined. Preliminary treatment is the removal of large items, sand and grit, floatables, and grease. Wastewater typically contains floatable materials—particularly fats, oils, and grease (FOG)—whereas (fresh) water does not. Removal is accomplished purely by physical processes such as screening and gravity, and is intended to protect the plant equipment. Primary treatment is the removal of suspended solids and organic matter, often by the addition of chemicals. Secondary treatment is a biological process that removes organics and suspended solids (and sometimes nitrogen and phosphorus), followed by disinfection. Tertiary treatment removes remaining suspended solids by fine filtering, and may include disinfection and nutrient removal. The standard today for wastewater is full secondary treatment, meaning that all the influent to a plant is given secondary treatment. When a plant cannot handle the flow, partial secondary treatment means that all the influent is given preliminary and primary treatment, but only part of it is given secondary treatment.

A typical process chain for secondary treatment is the activated sludge process shown in Fig. 3.[9] Large items are screened out. Dense noncontaminants are removed purely by gravity in the grit chamber. Primary treatment is a chemical/physical process, whereby small particles agglomerate (flocculate) and gravitate out. Secondary treatment is biological, in which microbes consume the dissolved and suspended organic matter. Disinfection kills most of the remaining contaminants.

Preliminary and primary treatment for wastewater is much like that for water. Preliminary screening of wastewater is necessary because large objects sometimes find their way into the sewer, an unfortunate example being construction debris illegally dumped into a manhole (alternatively, "maintenance hole"). (Note that a manhole cover is round for at least three reasons: (a) It will not fall through the hole, no matter how oriented; (b) it is easy to move by rolling; and (c) it need not be rotated to fit.) Primary treatment for both is nominally flash mix, flocculation, and sedimentation. Wastewater undergoes secondary treatment, in which microbes remove biological pollutants.

Both organic and inorganic particulates may be removed by settling, flotation, or filtering, depending on particle size and density. Carbon filtration is preferable to

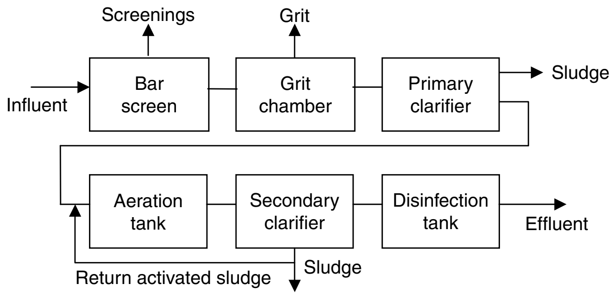


Fig. 3 Activated sludge wastewater treatment.

vaporizing, because the latter merely moves the substance from the water to the atmosphere.

Most reactions in waste treatment are of first order—that is, the rate of reaction is proportional to the concentration of the pollutant.

$$dC/dt = -k\ C$$

where C is concentration of pollutant (mg/L); t is time (min); k is reaction constant (mg/L-min).

Reactors are of three types: complete mix (or batch), plug flow, and dispersed flow. In a complete mix reactor, the reactor is filled; the reaction is allowed to take place; and then the reactor is emptied. The concentration of pollutant is equal throughout the tank. Complete mix reactors are approximately cubical and can be set up in sequence to provide an increasing proportion of removal. In a plug flow reactor, the flow moves through the reactor with the reaction taking place so that the concentration of pollutant is less at the outlet than at the inlet. These tanks are long in proportion to their length and width. Turbulence should be minimized to retain the form of the plug. Most reactors are of the dispersed flow type, intermediate between the other 2.[9]

Liquid and Solids Disposal

By the principle of conservation of mass, treatment does not make the contaminants disappear; it merely separates them from the water that bears them. When most of the contaminants have been removed and the water is sufficiently clean, it may be discharged to a natural waterway, such as a river, lake, or ocean. Most of the processes in Fig. 3 produce residuals. These impurities—such as sediment, sludge, waste washwater, and brine—are left behind to be treated and disposed of in other ways. Large items from screening and grit from the grit chamber are sent to landfill. Sludges from primary and secondary treatment are sent to a digester, which itself produces waste. Waste washwater from the filters is recycled, but because it adds to the throughput volume, it should be kept to a minimum.

Industrial Waste

For both direct dischargers and indirect dischargers, the content of industrial wastewater is regulated. Industrial users are allowed to send the first fraction of runoff from rainstorms to the sanitary sewer, for the purpose of keeping pollutants out of the storm sewers and, ultimately, out of natural waters after which it must be diverted to storm sewers.

Because POTWs are set up to treat organic waste, much industrial waste is incompatible with public treatment systems, so is usually subject to pretreatment. The objectives of such treatment are to prevent interference with the process in the POTW, to prevent pass-through of pollutants to the receiving waters, and to make possible reuse of the effluent and sludge from the POTW.[10]

Some treatment strategies are flow equalization to prevent shock loading to the POTW; solids removal by straining or settling; removal of FOG by dissolved air floatation or centrifuging; neutralization of high-pH or low-pH solutions; and hydroxide precipitation of heavy metals. A notable exception to the last is the removal of Cr^{+6}, which will not respond to hydroxide precipitation. Instead, it is converted by chemical reduction at low pH to Cr^{+3}, which can be removed by hydroxide precipitation. Dissolved inorganics may be removed by hydroxide precipitation, ion exchange, or membrane filtering.[10] It should be noted that diluting industrial wastewater to reduce the concentration of pollutants is not acceptable, and that "dilution of pollution is not a solution, and can lead to prosecution." The most important principle is segregation: keeping the pollutants separated so that they can be treated individually.

Issues

Wastewater treatment plants are designed to treat organic wastes. Other substances (nutrients in fertilizer, pharmaceuticals, etc.) can pass through and cause problems in receiving waters (e.g., algal growth and abnormal growth in fauna). Yet others, such as heavy metals and toxic chemicals, can cause interference (also called upset)—disruption of the process in the biological reactor.

Failed equipment can cause raw or partially treated wastewater to flow into storm drains and then into natural waters. Runoff from heavy storms can flow into sanitary sewers and overwhelm treatment plants. Cities with combined sewers are especially subject to this problem. Inadequately sized treatment plants will discharge partially treated wastewater in times of heavy flow.

Everything removed from wastewater must be disposed of. Sometimes, objections are raised to the release of volatiles into the atmosphere. Although the creators believe that digested sludge is a fertilizer rich in nutrients, others are not convinced.[11]

More radically, some have questioned the wisdom of the whole process of fouling great quantities of cleaned water and then cleaning it again.[12]

CONCLUSION

A basic requirement of human existence is an adequate supply of clean water. Today, very few have access to clean, untreated water. Wealthy societies obtain clean water by treating it, while poor ones often rely on polluted sources and suffer from the resulting waterborne diseases.

The idea that many municipalities draw from surface waters that are used for the disposal of wastewater is sobering, if not chilling. In the United States, the Safe Drinking Water Act mandates water treatment standards, and the Clean Water Act mandates wastewater treatment standards. Observance of these standards has made the practice of having a common source and sink acceptable.

The conventional water treatment process described in this article has been very effective in removing bacterial pathogens. Dechlorination to control disinfection byproducts created by chlorination was instituted as a result of the Safe Drinking Water Act. The most recent major issue is the resistance of *Cryptosporidium* oocysts to chlorination. Membrane filtration is an effective way to remove these and other very small suspended pathogens. New water treatment plants are likely to be based on this technique because of its effectiveness and ease of operation.[5]

The most important developments required in wastewater treatment are (a) building or expanding facilities to provide the capacity to subject all flow to complete secondary treatment, (b) repairing and maintaining the collection system to ensure that all wastewater reaches the treatment plant, and (c) finding practicable ways to dispose of the residual solids.

REFERENCES

1. Reichenberger, J. Unpublished notes. Loyala Marymount University: Los Angeles, 2005.
2. Jesperson, K. *Search for Clean Water Continues*. Available at: www.nesc.wvu.edu/ndwc (accessed 27 November 2006).
3. Judith Summers. *Broad Street Pump Outbreak*. Available at: www.ph.ucla.edu/epi/snow/broadstreetpump.html (accessed 27 November 2006).
4. United States Environmental Protection Agency, Office of Water. *The History of Drinking Water Treatment*. Available at: http://epa.gov/safewater/consumer/pdf/hist.pdf (accessed 28 November 2006).
5. Kawamura, S. *Integrated Operation and Design of Water Treatment Facilities*, 2d Ed.; Wiley: New York, 2000.

Was–Wat

6. Graber, C. *Desalination in Spain*. Available at: www.technologyreview.com/spain/water (accessed 27 November 2006).
7. Salt Lake City Department of Public Utilities. *Conduits of Civilization*. Available at: www.ci.slc.ut.us/utilities/NewsEvents/news1998/news9281998-1.htm (accessed 27 November 2006).
8. Greenberg, Arnold, et al., eds. *Standard Methods for the Examination of Water and Wastewater*, 21st Ed.; American Public Health Association (APHA), American Water Works Association (AWWA), and Water Environment Federation (WEF): Washington, D.C., 2005.
9. Tchobanoglous, G.; Burton, F.L.; Stensel, H.D. *Wastewater Engineering, Treatment and Reuse*, 4th Ed.; Metcalf & Eddy Inc., McGraw-Hill: New York, 2003.
10. Water Environment Federation Pretreatment of Industrial Wastes In *Manual of Practice FD-3*; Water Environment Federation: Alexandria, VA, 1994.
11. Voters seek to block sludge, *Los Angeles Times* 2006, January 2. L.A. Fights Kern County Sludge Ban. *Los Angles Times* 2006, August 16.
12. Stauber, J.C.; Sheldon, R.S. *A Brief History of Slime*. Available at: www.prwatch.org/prwissues/1995Q3/slime.html (accessed 27 November 2006).

Water Source Heat Pump for Modular Classrooms☆

Andrew R. Forrest
Government Communications Systems Division, Harris Corporation, Melbourne, Florida, U.S.A.

James W. Leach
Industrial Assessment Center, Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, North Carolina, U.S.A.

Abstract

This work investigates and recommends design improvements for a water source heat pump system for mobile homes and modular classrooms. It builds on a previous study that tested a 3-ton geothermal heat pump in a modular classroom at Wilson Mills Elementary School in Johnston County, North Carolina. Water stored in flexible plastic bladders resting on the ground underneath the classroom served as the heat source. The bladders were filled with 2000 gal of saltwater. Using Transient Systems Simulation Program (TRNSYS), a model of the original system was constructed and validated by comparing model predictions with measured performance. Transient systems simulation program models of several new designs were constructed to evaluate potential design improvements. The system models were evaluated based on predicted performance for a typical meteorological year, and on other criteria, such as initial cost, maintenance, and portability. This resulted in a new optimized system design in which the water storage volume is reduced to 120 gal, and the predicted electrical energy requirements are about two-thirds of those of an air source heat pump. The predominant design improvement is to replace the bladders with heat exchangers constructed of PVC pipe. Design, costs, and assembly procedures for the PVC heat exchanger are presented in this study.

INTRODUCTION

The purpose of this project is to optimize the operating parameters and design of an earth-coupled heat pump system with aboveground water storage.[1] A computer model was created in a Transient Systems Simulation Program (TRNSYS) to predict the best geothermal heat pump design. Available experimental results from the previous project were compared with predicted results to validate the model. Afterwards, the computer model was used to predict how changes in system parameters would affect the energy use of the water source heat pump.

Background

Two previous studies were performed to develop geothermal heat pump technology specifically for mobile classrooms, homes, and offices. Researchers at Progress Energy Carolinas (formally Carolina Power & Light) conceived the original design and funded the first study, which was conducted by a graduate student at North Carolina State University.[2] A 1-ton water source heat pump was installed on a mobile office building, and the water in the system circulated through a 7000-gal plastic bladder that rested on top of the ground. The bladder was insulated from the atmosphere to provide some freeze protection, and heat strips were installed in the bladder to heat the water if temperatures approached the freezing point.

The second study, which was funded by the North Carolina State Energy Office, was twofold, containing a theoretical analysis and experimental validation of a low cost version of the original design. The second heat pump was installed in a modular classroom in Johnston County, North Carolina at Wilson Mills Elementary School (WMES). The heat pump capacity increased from 1 ton in the original mobile office to 3 ton for the modular classroom, and the volume of the storage bladders decreased from 7000 to 2000 gal. The bladders underneath the classroom were not insulated, and saltwater was used to prevent freezing. Fig. 1 is a schematic of the water source heat pump system at WMES. A photograph of the classroom with the heat pump installed is shown in Fig. 2. The theoretical study[3] predicted that the geothermal heat pump would use approximately one-half the energy that a similar air source heat pump would use for the same classroom.

The experimental part of the study[4,5] compared the performance of the geothermal heat pump system with that

☆ This entry originally appeared as "Water Source Heat Pump for Modular Classrooms" in *Energy Engineering*, Vol. 102, No. 2, 2005. Reprinted with permission from AEE/Fairmont Press.

Keywords: Water source heat pump; Modular classrooms; Mobile homes; TRNSYS; PVC.

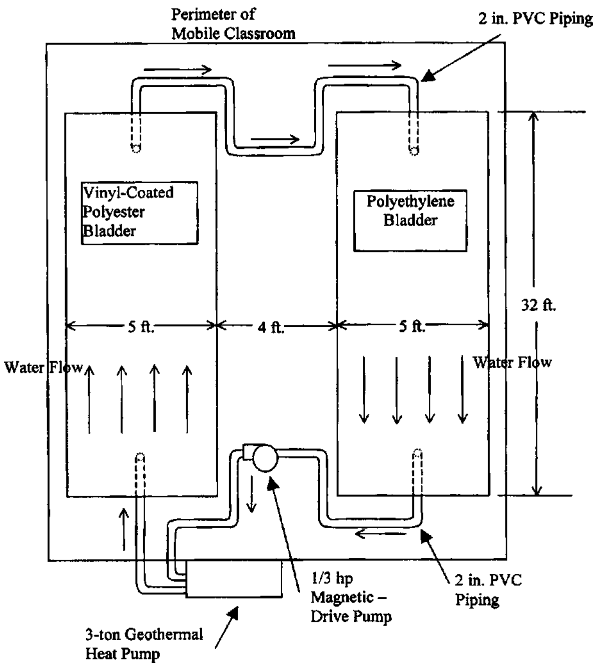


Fig. 1 Layout of original bladder system.
Source: From North Carolina State University and ASHRAE (see Refs. 4 and 5).

of a nearly identical air source heat pump at an adjacent classroom. The experimental results showed that the heat pump did use about one-half as much electricity as the air source heat pump while in heating mode. However, in cooling mode, the geothermal heat pump used about 80% as much energy as the air source heat pump. Subsequent study showed that the ventilation rate in the classroom with the air source heat pump was well below the value recommended by ASHRAE. This changed the building load in favor of the air source heat pump. In addition, the air source heat pump could not maintain the interior of the classroom at comfortable temperature during the hot summer months.

Fig. 2 Modular classroom style where experiments are conducted.
Source: From North Carolina State University and ASHRAE (see Refs. 4 and 5).

The purpose of the current work is to improve the design of the water source heat pump so that it is more energy efficient and reliable. Also, a reduction in the initial cost of the heat pump system and additional system mobility is desired. The work will also demonstrate the performance advantages of the improved design by comparing two water source heat pumps with three air source heat pumps in modular classrooms.

In the first phase of this work, candidate design improvements were evaluated, new heat exchangers were designed and fabricated, and a test site was selected. A TRNSYS model of the classroom and heat pump was developed. The model was calibrated by comparing predicted performance with experimental data from WMES. Wake County Public School Systems agreed to allow their mobile classrooms to be used to demonstrate the water source heat pumps for a period of 1 year. Several potential test sites were considered, and Davis Drive Elementary School (DDES) was selected because the site contained multiple school trailers with similar building loads. Several data loggers were left at selected trailers to confirm this assertion. Measurements from the crawl space from two of these trailers were used to design the PVC heat exchangers.

The second phase of this study is in progress. The PVC heat exchangers and water source heat pump were installed in June 2003 as shown in Fig. 3. Data are being collected to determine the system's actual economic and physical performance. The data will be compared with predicted performance to further validate the TRNSYS model. The design and volume capacity of the bladder are substantially different from those of the PVC heat exchanger, which is evident in Figs. 3 and 4.

Previous Project Problems

The experiment at WMES provided valuable experimental results to validate previous theories. However, several problems with the water source heat pump design became evident as the experiment was being conducted. Initially,

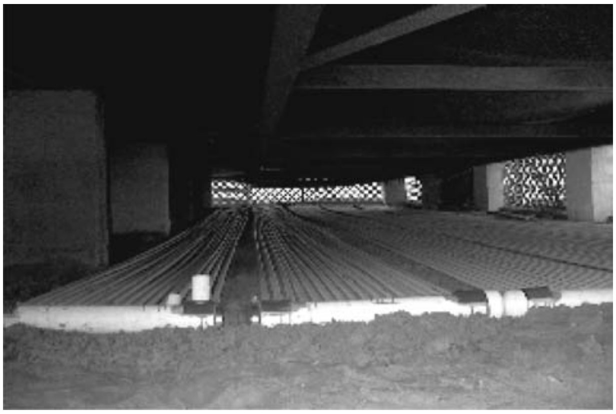


Fig. 3 PVC heat exchangers installed at DDES.

Was–Wat

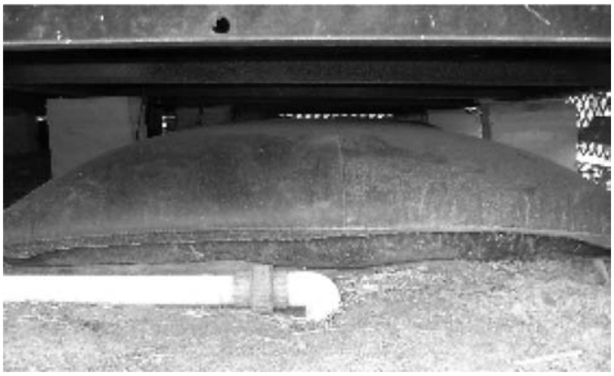


Fig. 4 Installed bladder underneath modular classroom at WMES.
Source: From North Carolina State University and ASHRAE (see Refs. 4 and 5).

one of the bladders split at the seam during filling, and required replacement. In addition, the saltwater used for freeze protection was corrosive to the pump and heat pump. The volume of the brine solution (2000 gal) presented disposal and portability problems.

The original bladder that split at the seam was made from a high-density polyethylene material, and cost $250. The bladder that was used as a replacement was constructed of a vinyl coated woven polyester fabric, and cost $750. The $250 polyethylene bladder that did not fail is shown filled with 1000 gal of saltwater in Fig. 4.

With a 2000-gal storage volume, saltwater provided a low-cost option when compared with using antifreeze for freeze protection. However, the saltwater corroded the direct drive pump and required the use of a more expensive, magnetic drive pump. Furthermore, the saltwater destroyed the soft solder copper pipe connections. The leakage of saltwater around the connections caused the base of the heat pump enclosure to rust.

Two thousand gallons of saltwater cannot be easily transported or easily disposed of when the modular classroom is relocated to another site. To ensure freeze protection, salt was added to the water in the bladders to a 15% solution.[4] This volume of saltwater would present a disposal problem at the end of the equipment life.

The North Carolina Energy Office recognizes the need for efficient and economical water source heat pumps suitable for mobile buildings. Therefore, the present work was funded to further develop and refine the system design. One potential design improvement would incorporate low cost solar collectors to heat the storage water.

MODEL EXPLANATION AND VALIDATION

TRNSYS Description

The TRNSYS was chosen as the software to model the geothermal heat pump system for three reasons. Transient systems simulation program is a Fortran language-based program, which allowed for a ground temperature distribution finite difference model and an earth-coupled heat exchanger to be modeled in conjunction with the existing TRNSYS components. Transient systems simulation program is well known for its ability to model solar radiation and solar collectors in thermal systems. The software was used to predict the performance of solar collectors in the geothermal heat pump system. Since the system being modeled in TRNSYS consists of multiple components, it allowed a large, complex system to be broken into smaller parts. Thus, it was convenient to analyze the effects of modifying various parameters in the system.

WMES Experimental Results and Model Predictions

It was necessary to construct a TRNSYS model of the previous experiment to validate the calculations for the new design. The model predicts the heat pump electrical energy requirements based on typical meteorological year (TMY) data. Unknown parameters in the model, such as the heat transfer coefficient between the bladder surface and the ambient air, were adjusted until it predicted most accurately the performance of the previous experiment. The theoretical data and experimental results presented in Fig. 5 are organized to correspond directly to the days that actual heat pump energy use was recorded. The model correlates very closely with recorded data during the spring and fall months. When divergences do occur, the majority of the differences in experimental and theoretical energy use can be attributed to differences in actual degree days and TMY degree days as shown in Fig. 6. The two exceptions occur during the summer months. In the summer 2000, the degree days difference accounts for some of the divergence, while the remainder could be attributed to larger infiltration and internal loads due to increased student activity levels. In summer 2001, the classroom was not used for summer school, as it was the previous summer. The disparity in energy use in the winter months can be explained by differences in degree days. The predicted annual energy use was 9337 kWh, and the actual annual energy use during the previous study totaled 10,783 kWh. This underestimation should be taken into account when viewing the subsequent model predictions for the new designs.

NEW DESIGNS

The initial results from the TRNSYS models showed that the heat pump performance is not very sensitive to the volume of water stored, but is sensitive to the surface area available for heat transfer from the water to the ground and the ambient air. This led us to consider the possibility of replacing the bladders with PVC heat exchangers as shown

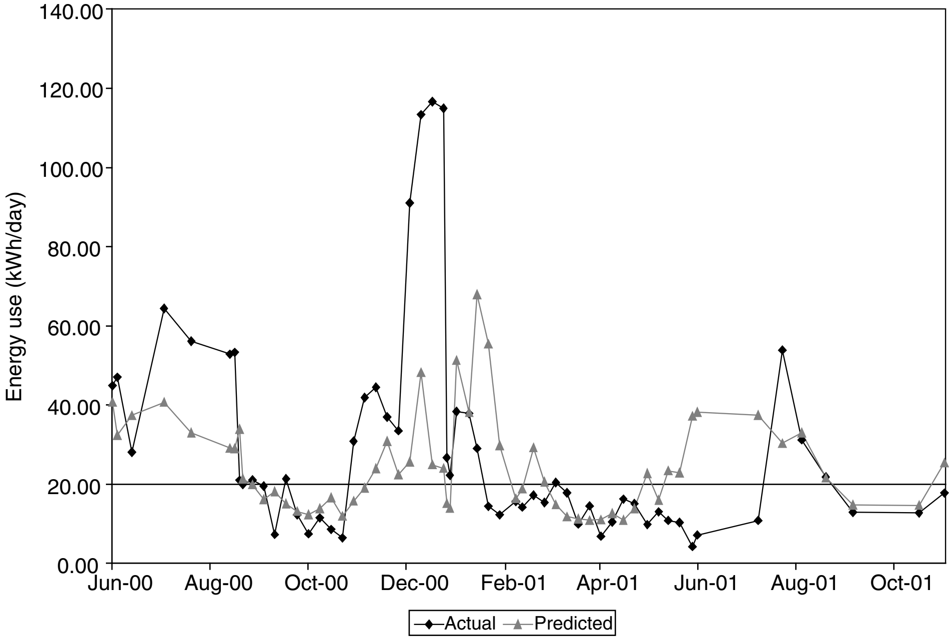


Fig. 5 Energy use comparison for original geothermal heat pump study.

in Fig. 3. Heat exchangers such as this would be more durable than the bladders and would be easier to move. Also the surface area and volume can be optimized independently by changing the diameter of the PVC pipe. The surface area-to-volume ratio for the small diameter pipe in Fig. 3 is much greater than the corresponding ratio for the bladders.

Surface Area and Volume Effects

Parametric analyses were performed to establish the dependence of the annual electrical energy requirements on the volume of the water storage and the surface area available for heat transfer. The heat transfer coefficient for convection to the ambient air was assumed to be constant and independent of pipe diameter. The surface area available for heat transfer to the ground was assumed to be 30% of the total surface area. A simulation with hourly time steps was performed using TMY data.

The trends predicted in Fig. 7 are the fundamental reason for using a heat exchanger design in place of the bladders. The predicted annual energy use of the heat pump in the bladder system (with two 1000 gal bladders and

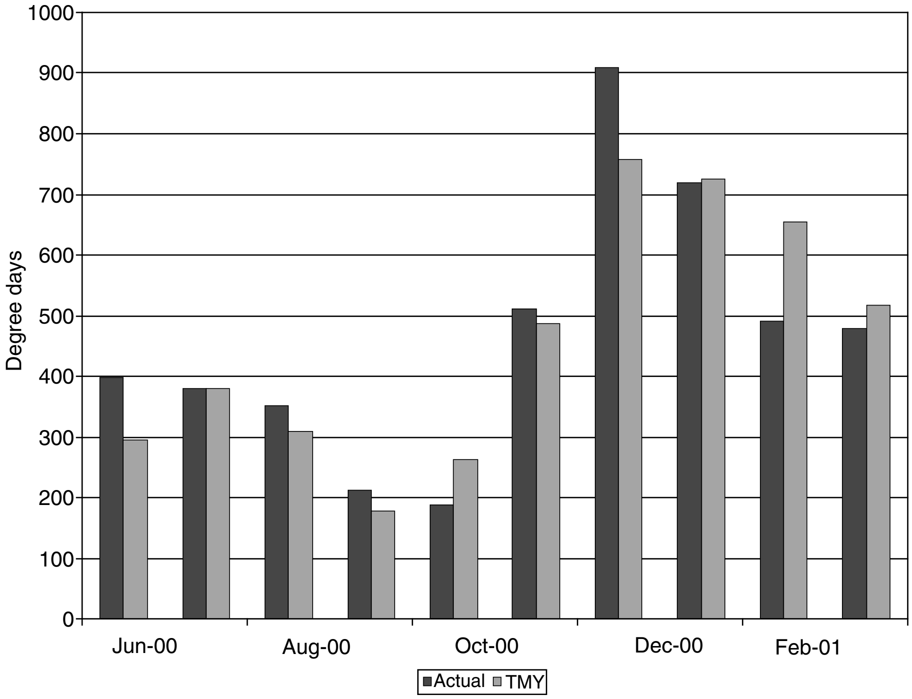


Fig. 6 Degree day comparison for portion of original geothermal heat pump study.

Was–Wat

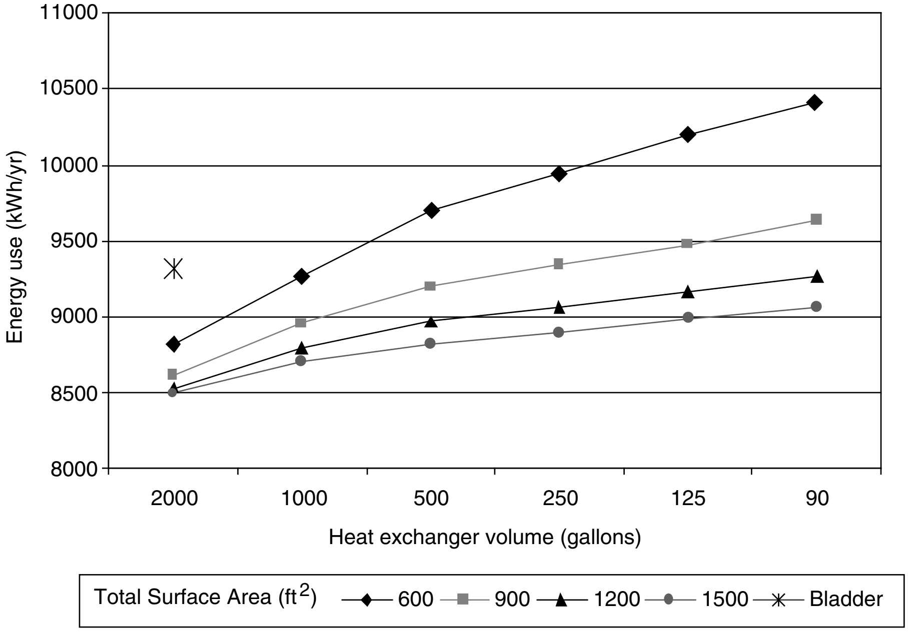


Fig. 7 Water source heat pump and pump annual energy use for various surface areas and volumes.

640 ft^2 of total surface area) is included for comparison. The difference in energy use is due to different percentages of surface area allocated for heat transfer to air and water between the bladder and heat exchanger models. Furthermore, the theoretical heat transfer coefficient for the heat exchangers to air is twice as large as the heat transfer coefficient for the bladders to air. The model predicts that a heat pump coupled with a heat exchanger having a volume of 120 gal and a surface area of 1200 ft^2 would require less electricity than one coupled with 2000 gal of bladder storage. The 120 gal storage volume is much more desirable for mobile buildings.

As shown in Figs. 8 and 9, the fluid temperature in heat exchangers with a volume of 120 gal and a surface area of 1200 ft^2 is fairly close to the ambient air temperature. The fluid in the bladders is warmer in the summer and cooler in the winter than the fluid in the heat exchangers. Therefore, the coefficient of performance (COP) of the heat pump is expected to be greater when the PVC heat exchangers are used. The average amplitude of daily temperature fluctuation in the bladders and PVC heat exchangers was predicted to be about 10 and 20°F, respectively.

Solar Collector Investigation

One of the objectives of this work is to evaluate the economics of using low-cost unglazed solar collectors to improve the heat pump coefficient of performance. The storage water would flow through the collectors before it arrives at the heat pump. The storage water, which is

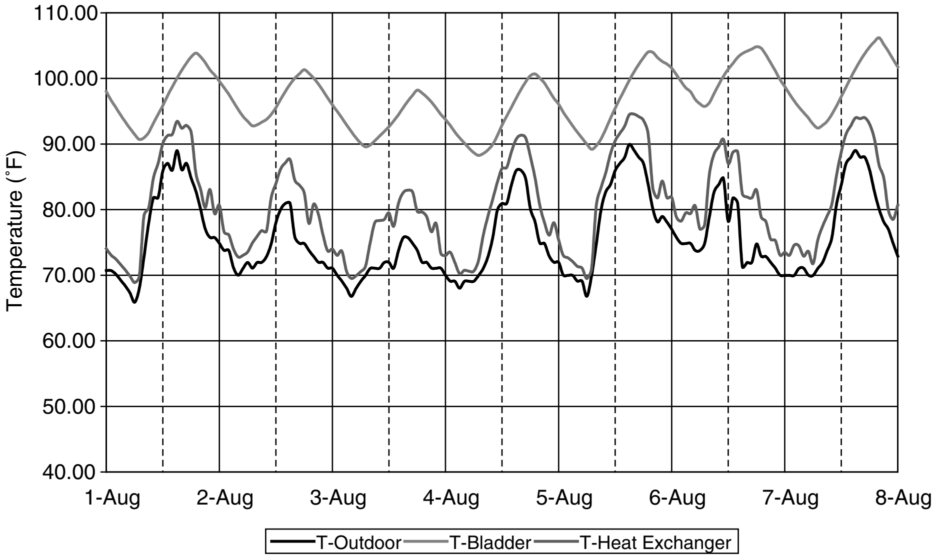


Fig. 8 Typical predicted summer temperatures.

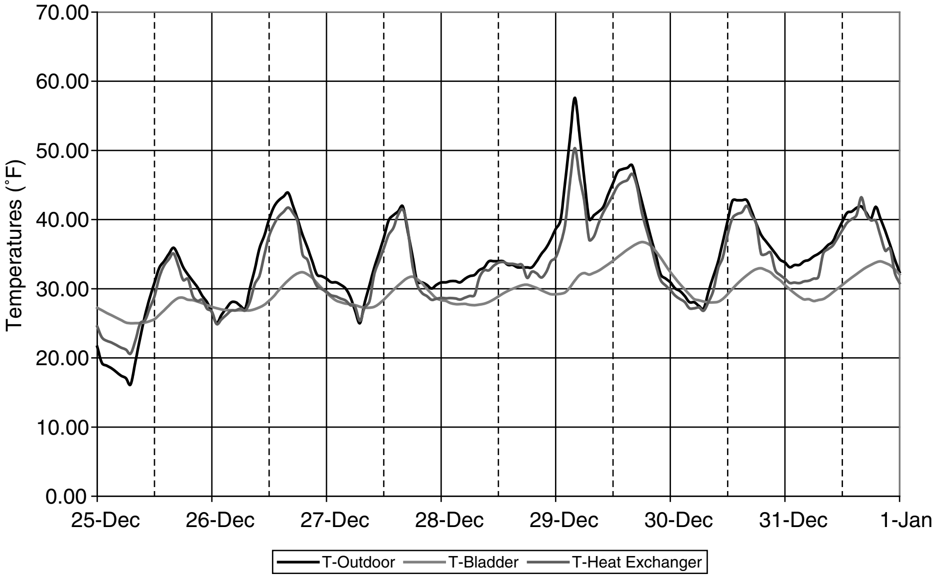


Fig. 9 Typical predicted winter temperatures.

always colder than the ambient temperature in the winter, would be heated by the sun and by convection from the air. During the summer, the collectors could be shaded so that the storage water is cooled by convection.

The interest in modeling solar collectors as an energy savings measure motivated the decision to use TRNSYS. However, the TRNSYS code also proved to be useful for modeling systems without solar collectors. The predicted electrical energy savings yielded by the solar collectors was much less than expected. A system with three 4×8 ft unglazed collectors was modeled. The collectors were plumbed in parallel to prevent excessive pressure drop. The model predicts that solar collectors provide more savings for a system with bladders than for a system with PVC heat exchangers. Still, the system with solar collectors and PVC heat exchangers uses less energy. The larger storage capacity of the bladders enabled them to store the energy gained by the solar collectors. Fig. 10 shows the energy savings from using solar collectors with bladders and heat exchangers. Despite the increased energy savings from using the solar collectors with the bladders as opposed to heat exchangers, the predicted payback period for the three solar collectors being modeled would exceed 10 years. Further evidence of the solar collectors ineffectiveness can be seen by the fraction of time the fluid flows through the collectors. The simplest control strategy causes fluid to flow though the collectors when the sun is shining and the heat pump is heating, or when the sun is not shining and the heat pump is cooling. In this case, the solar collectors are used only about 15% of the time. More complex strategies increase the time that the collectors can be used, but do not have much effect on energy usage.

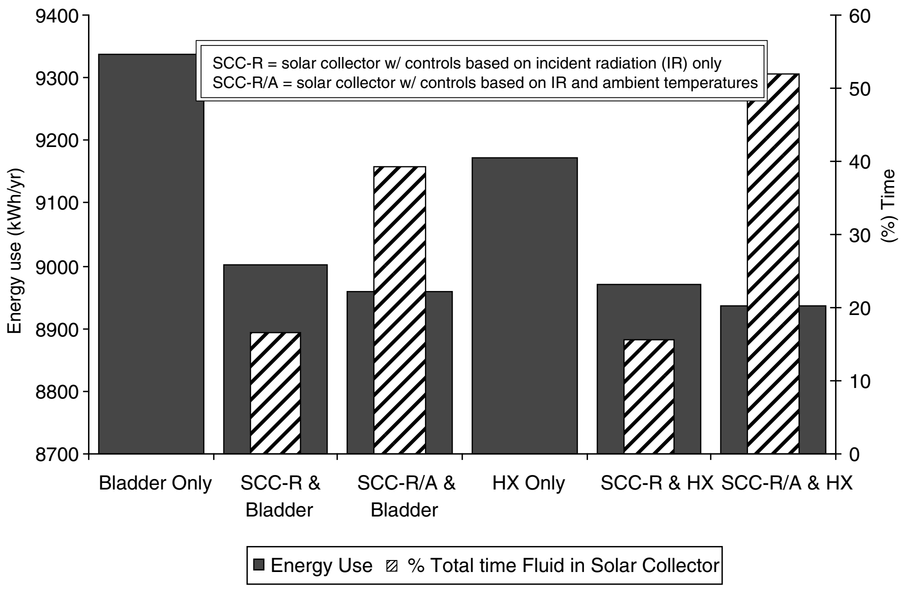


Fig. 10 Geothermal heat pump energy use with solar collectors.

Another issue to be examined is the optimum orientation of the solar collectors. For appearance, it is desirable to have the collectors lie flat on the roof. Measurements were taken at DDES to determine the slope of the classroom roof, β, and the azimuth, γ, of the most southern-facing roof. β was measured to be about 14° and γ was measured to be about 30° east of south. The optimum case for solar collector orientation was determined to be at a slope of 47° and facing south. According to the Florida Solar Energy Center, a solar collector performs best facing south, but a 30° variation in either direction will still allow the solar collector to capture 90% of the maximum solar energy available.[6] The recommended optimum slope that a solar collector should be positioned for winter use is the location's latitude plus 15°. In Raleigh, North Carolina, this would equate to about 51°. The TRNSYS model predicted that energy savings associated with changing the orientation of the collectors are very small. Thus, the solar collectors could lie flat on the roof at DDES.

Fluid Flow Effects

Parametric analyses were performed to determine whether an optimum flow rate exists, and to determine whether it coincides with the heat pump manufacturer's recommended flow rate of 7 gpm. The simulation, which was carried out over the range of flow rates, accounts for improvements in heat pump COP and energy costs to drive the circulation pump. The optimum flow rate for heat pumps with solar collectors and heat pumps without solar collectors was determined to be 13 gpm, which differs from the manufacturer's recommended flow rate, but is not excessive. As shown in Fig. 11, the differences in energy consumption between a system with and without solar collectors dwindle as the flow rates are increased.

Three different types of fluids were modeled for the same heat pump: pure water, 25% ethylene glycol, and 25% brine solution. The system containing the brine solution properties performed the worst. The system containing the ethylene glycol solution yielded about a half percent energy savings when compared with the brine solution. The system containing pure water performed the best (about one-third percent savings over the ethylene glycol solution). The results of the simulations were in line with expectations due to the variations in specific heat of the three solutions.

Thermostat Setback

Another objective of this work is to estimate the energy savings made possible by automated thermostat setback. All mobile classrooms in Wake County, North Carolina are equipped with radio-controlled night setback controls. During the winter months, the thermostat is supposed to be set back to 55°F typically from about 4:00 P.M. to 6:00 A.M. In the summer months, the heat pump is supposed to be turned off during a similar time frame. Some of the night setback thermostat controls were not functioning during our visits to the classrooms. In addition, these controls can be manually overridden with the thermostat inside the classroom. The night setback feature was modeled in TRNSYS to predict the energy savings generated by the use of these controls for a geothermal heat pump system with PVC heat exchangers and for an air source heat pump. The results generated by the TRNSYS model, which are summarized in Fig. 12, show that the night setback controls reduce electrical energy consumption by about 30%.

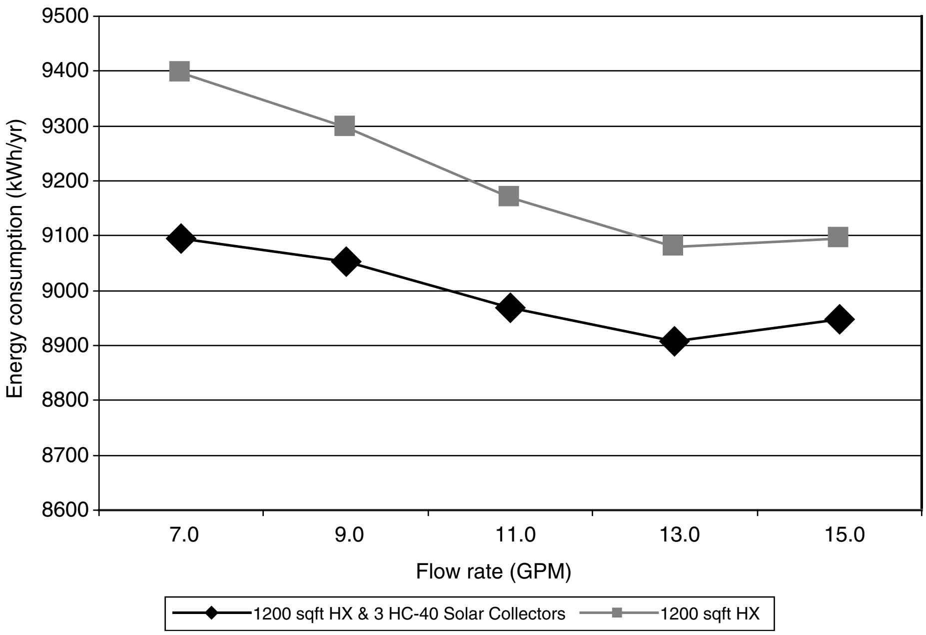


Fig. 11 Heat pump annual energy use for manufacturer's flow rate range.

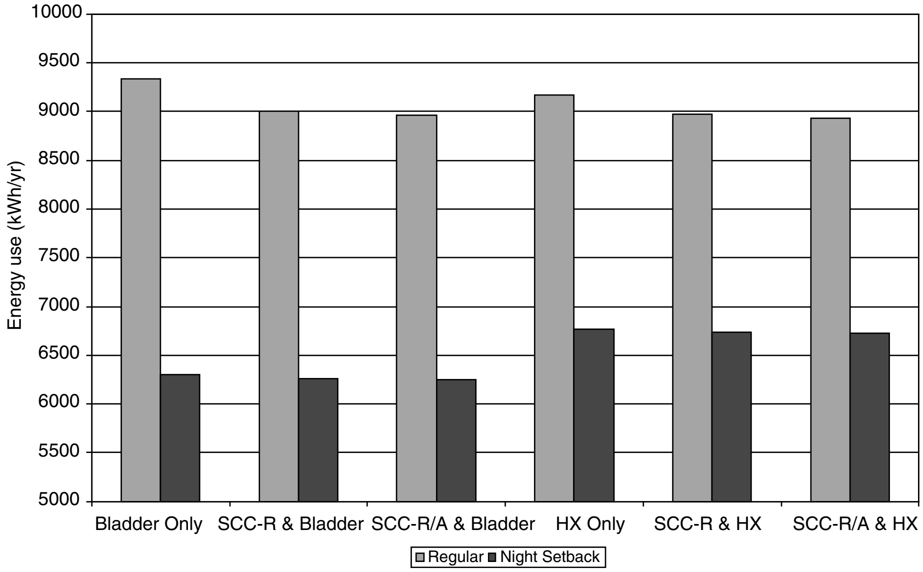


Fig. 12 Energy use with night setback controls.

Optimum System

An optimum system was established based on the analysis presented in the previous sections and on the existing layout of the concrete piers in the classroom crawlspace. The optimum system was determined to have a surface area of 1200 ft^2 and a volume of 120 gal. The flow rate of the ethylene glycol solution in the system was prescribed to be 13 gpm. Fig. 13 compares the daily energy use of the original bladder system and the optimum PVC heat exchanger system. The total predicted electrical energy consumption of the optimized heat pump system was 9069 kWh/yr, a predicted energy savings of 3% compared with the original system. The optimum system used much less energy in the summer months, but used slightly more or about the same throughout the rest of the year.

PVC HEAT EXCHANGER DESIGN

The PVC heat exchangers will be more durable and less likely to leak than the bladders were. The significant reduction in volume of the PVC heat exchangers will allow an ethylene glycol solution to be used instead of a brine solution, eliminating the problems caused by saltwater. A less-expensive, direct drive pump can be used to circulate the ethylene glycol solution. In addition, the ethylene glycol solution can be reused when the mobile classrooms are moved, further reducing the overall cost of the system.

The heat exchanger that evolved as a result of this work is made from PVC pipe, and is comprised of a header and footer, and 0.5-in. PVC pipe (Fig. 14). Space limitations in the crawl space of mobile buildings influenced the design

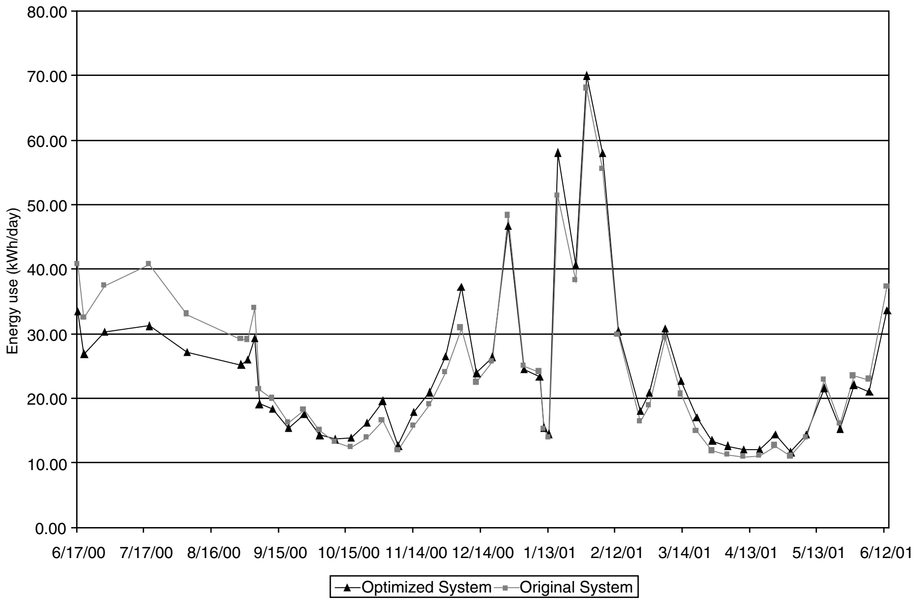


Fig. 13 Annual energy comparison of original and optimized systems.

Was–Wat

of the heat exchanger. The heat exchanger was designed to be approximately 2-ft wide, and to run the length of the space it is intended to occupy. The heat exchangers were designed to this width to ease manufacturing and handling of the headers, and to maintain the desired fluid velocity of about 1 ft/s. An additional benefit from using the multiple 2-ft wide heat exchangers would be realized if a replacement were required. The heat exchangers are connected in series with a hose and a clamp fitting.

The available area for heat exchanger placement varies between mobile classrooms. Fig. 15 shows a layout for a typical classroom. This style layout was used at DDES. The piers under some mobile buildings are arranged in lines that extend across the width of the building instead of the length. In this case, the heat exchanger should also extend across the width to fully utilize the available space.

Pressure Drop

The heat exchanger was designed to have minimal pressure drop. The diameter of the headers ensured that the pressure drop in each header could be neglected during the overall pressure drop calculations. Based on 15 or 16 tubes per heat exchanger and an optimum system flow rate of 13 gpm, the velocity through each of the tubes in the heat exchangers is about 0.75 ft/s. The total pressure drop across ten heat exchangers 30-ft long is predicted to be less than 2 ft of water. By comparison, the pressure drop across the water to refrigerant heat exchanger in the heat pump is about 20 ft. A one-third horsepower centrifugal pump will provide the required flow of 13 gpm.

Material Specification

The heat exchanger header shown in Fig. 14 is made of 2.5-in. SCH-40 PVC pipe and 3-in. SCH-80 reinforcement where holes for the risers are to be drilled. The reinforcement is made of quarter sections of the 3-in. SCH-80 PVC pipe. The reinforcement was added to

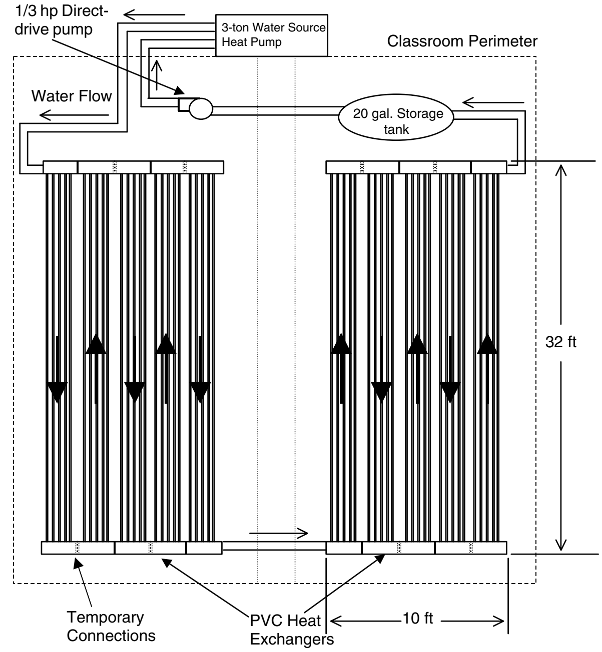


Fig. 15 Layout of PVC heat exchanger system.

increase bonding surface area between the headers and the risers. By adding the reinforcement, the wall thickness was increased by approximately 60%. The inner diameter of the 3-in. pipe is approximately equal to the outer diameter of the 2.5-in. pipe, which ensures a natural fit. This extra step significantly decreases the likelihood that the heat exchanger will leak at the riser holes.

The tubes of the heat exchanger were made from 0.5-in. SDR-13.5 PVC pipe. This thin wall pipe was selected rather than 0.5-in. SCH-40 for its lower cost, lower thermal resistance to heat transfer, and greater volume capacity. The result of using SDR-13.5 PVC instead of SCH-40 PVC yielded a cost savings of 30% and a volume capacity increase of 33%.

An additional 1-in. hole was drilled in the header assemblies so that an adapter could be installed to allow air to be purged from the system. Emersion thermocouples were installed at theses locations so that temperatures could be monitored throughout the network of heat exchangers.

The risers from each header are joined by standard SCH-40 0.5-in. couplings and standard 10 or 20 ft lengths of 0.5-in. pipe. The 20 ft lengths contain a belled end, which reduce the number of couplings required. The use of standard PVC lengths and couplings simplifies the assembly of the heat exchangers during installation, and helps to keep costs low.

The headers are connected to each other via 3-in. lengths of nitrile (Buna-N) hose and stainless steel worm gear hose clamps. Nitrile (Buna-N) hose was chosen because it will not deteriorate under the temperature range expected in the system (20°F–120°F). The hose has an "A"

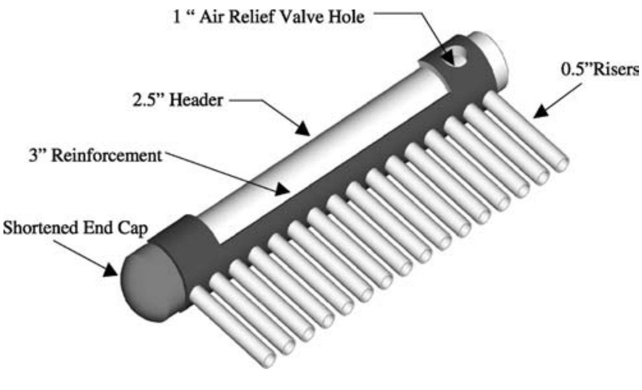


Fig. 14 PVC heat exchanger header design.

rating for transporting ethylene glycol (antifreeze) solutions.[7] Because the water pressure in the heat exchangers is relatively low, hose connections can be used in place of more expensive PVC unions.

Materials and Labor Cost

North Carolina State University's overall cost to manufacture and install the 1200 ft^2 surface area PVC heat exchangers in one classroom was approximately \$1000. This cost represents retail material costs, and a \$10/h labor rate. The cost was approximately \$500 less than the uninstalled cost of two vinyl-coated polyester bladders. The less expensive and less reliable bladders made from polyethylene cost about \$500 less than the PVC heat exchangers.

CONCLUSIONS AND RECOMMENDATIONS

A TRNSYS computer model of a mobile classroom with a water source heat pump was constructed to predict the effects of proposed design modifications. Comparisons of predicted performance with experimental data from a previous project were used to validate the computer model. The computer model was then used to evaluate proposed design modifications. Parametric analyses showed that solar collectors are not cost effective in this system. However, the TRNSYS model made it possible to design PVC heat exchangers to replace large bladders employed in previous designs. The heat exchangers are expected to improve energy efficiency, reduce initial costs, increase mobility (in the event the system is moved), and improve reliability.

The predicted and actual water source heat pump electrical energy use correlated as well as could be expected using actual data and TMY data. Thus, the predictions and conclusions presented in this study should provide significant insight into the actual behavior of the enhanced geothermal heat pump system. The redesigned geothermal heat pump system is currently being tested at DDES and will be monitored for a test period of 1 year.

ACKNOWLEDGMENTS

The authors wish to thank the North Carolina State Energy Office for its technical guidance and financial support. We especially acknowledge the help of our project monitor, Mr Bob Leker, who worked many hours with us to ensure that our analyses are sound, and that our objectives are in line with those of the State Energy Office.

REFERENCES

1. Forrest, A.R. *Optimization of a Geothermal Heat Pump System with Aboveground Water Storage*; North Carolina State University: Raleigh, NC, 2003; [master's thesis].
2. Rich, J.L. *Thermal Performance of a Bladder Tank Water Storage System Coupled in a Closed Loop Series Configuration with a Water Source Heat Pump*; North Carolina State University: Raleigh, NC, 2003; [master's thesis].
3. Smith, K.M. *Geothermal Heat Pumps and Modular Classroom Units*; North Carolina State University: Raleigh, NC, 1999; [master's thesis].
4. Soderberg, E.W. *Test of an Earth-Coupled Heat Pump With Aboveground Water Storage*; North Carolina State University: Raleigh, NC, 2001; [master's thesis].
5. Soderberg, E.W.; Leach, J.W. Test of an earth-coupled heat pump with aboveground water Storage. *ASHRAE Transactions Symposia*; CI-01-7-2; ASHRAE; 600–607.
6. Florida Solar Energy Center. Available at: http://www.fsec.ucf.edu/en/consumer/solar_hot_water/pools/q_and_a/#Orientation (accessed on May 2003).
7. Arimes, T. *HVAC and Chemical Resistance Handbook for the Engineer and Architect*; BCT: Lexington, KY, 1994.

Water-Augmented Gas Turbine Power Cycles

Ronald L. Klaus
VAST Power Systems, Elkhart, Indiana, U.S.A.

Abstract
Conventional gas turbine systems convert only a small fraction (25%–35%) of the available chemical energy in their fuel into useful mechanical or electrical energy. One of the most common ways to improve these power cycles is through the imaginative use of water. Such cycles fall into four families: (1) combined cycles (CCs), which convert some otherwise wasted energy into steam to drive a steam turbine, (2) cycles that inject water into the compressor train and thus reduce the work of compression, (3) cycles that humidify air before combustion in a very thermodynamically efficient manner, and (4) cycles that inject steam or water into the combustor or immediately thereafter. All of these approaches are discussed and certain performance parameters are presented.

NOMENCLATURE

- R Universal gas constant [(energy/mole * temperature)]
- T_C absolute temperature of cold reservoir (^{0}K or ^{0}R)
- T_H absolute temperature of hot reservoir (^{0}K or ^{0}R)
- T_o absolute base or environmental temperature (^{0}K or ^{0}R)
- TIT turbine inlet temperature (any temperature units)
- W work per mole (energy/mole)
- γ ratio of constant pressure to constant volume heat capacities
- υ isentropic efficiency

INTRODUCTION

Conventional gas turbines convert only a small fraction (25%–35%) of the chemical energy in a typical fuel to usable mechanical or electric power. Because of the second law of thermodynamics, not all of that energy can be converted, even with perfect machinery. However, there has been an ongoing effort to develop gas turbine systems that come closer to the maximum second-law efficiencies. One approach has been to use water in imaginative ways to boost thermodynamic efficiency. One of the most comprehensive recent surveys of the literature that is worthy of special mention was done by Maria Jonsson.[1]

Water-augmented cycles fall into four categories: (1) cycles that recover otherwise wasted energy from the exhaust stream and use it to produce steam to power a steam turbine, (2) cycles that inject water into the compressor train to reduce the work of compression and therefore make more of the expander work available for useful purposes, (3) cycles that evaporate water into the air stream, which adds mass to be expanded and also permits better recovery of wasted exhaust gas energy, and (4) cycles that recover heat from the exhaust gas and use it to produce steam or hot water that is injected back into the combustor. Combinations of these ideas have also been proposed. Besides making electric power, each of these kinds of cycles has its own particular capacity to recover heat in the form of hot water and steam, which can be used for heating purposes and also for refrigeration.

Obtaining high efficiency of the energy conversion is only one side of the design problem. In general, the more efficient the energy conversion process, the higher is its capital cost. Thus, the best design has to balance these two issues and produce a power cycle that has the greatest thermoeconomic potential. Optimum design is also affected by plant capacity. Because capital costs are proportional to plant capacity raised approximately to the 0.8 power, larger plants favor higher capital cost expenditures to obtain greater efficiency, whereas the opposite is true for smaller systems.

The economic analysis of these cycles goes beyond the scope of this survey. However, one measure of the capital cost is the specific power, defined as the net power produced divided by the total mass flow through the turbine. The higher this mass flow—and thus the smaller the specific power—the larger the major equipment that handles it and thus the larger its capital cost. Thus, specific power is a rough but useful measure of the capital cost of the power cycle.

BACKGROUND

Carnot Efficiency

One way of quantifying the second-law limitation on the conversion of energy to work is through the so-called

Keywords: Gas turbines; Wet cycles; Recuperation; Brayton cycle; Combined cycle; Compressor cooling; RWI cycle; HAT cycle; STIG cycle; Combustion; Combined heat and power.

Table 1 Carnot efficiencies

T (°F)	Carnot efficiency (%)
1600	74.8
1700	76.0
1800	77.0
1900	78.0
2000	78.9
2100	79.7
2200	80.5
2300	81.2
2400	81.9
2500	82.5

Carnot efficiency.[2] The Carnot efficiency can be calculated to be:

$$\text{Carnot efficiency} = \frac{T_H - T_C}{T_H}$$

Table 1 shows efficiencies for the range of combustion temperatures typical in present gas turbines. These efficiencies are quite high and show that there is considerable room for improvement in efficiency over the typical 25%–35% efficiencies of conventional gas turbines. Notice also that the higher the combustion temperature, the higher the potential efficiency.

Computer Systems for Cycle Calculations

Many computer programs have been developed to analyze various power cycles.[3] Two of the most extensive commercial systems that are specially oriented toward power systems are Simtech's IPSEPro[4] and Thermoflow.[5] The simulations presented here were performed with the latter system. These and systems like them are placing increasing emphasis on economic as well as thermodynamic analysis of these cycles.

Choice of Machines for Simulation

Rather than studying specific machines, the bulk of this entry will deal with a range of "rubber machines" whose pressure ratios and turbine inlet temperatures (TITs) are taken over ranges typical for actual gas turbine systems (see Table 2, which shows ranges of TITs as a function of pressure ratio for a sample of typical gas turbines). For the purposes of this article, results will be presented for three representative machines whose pressure ratios, TITs, and power levels are also shown in that table.

Blade coolant flow is usually considered proprietary information and is difficult to estimate. This entry will use the following empirical equation for blade cooling as a percentage of the main flow entering the expander,

Table 2 Pressure ratios and TIT's of machines investigated

Pressure ratio	Low TIT (°F)	High TIT (°F)	Number of machines
5	1600	1700	2
10	1700	2100	5
15	1900	2500	7
20	2100	2500	5
25	2200	2300	2
30	2100	2300	3
35	2100	2400	4
		Total number	28

Base machines

Pressure ratio	TIT (°F)	Net power (kW)
10	1900	4600
20	2200	25500
30	2399	191000

$$\text{Blade cooling } \% = 0.0013(\text{TIT})^{1.286}$$

where TIT is in °C. This equation is based on cooling flow data for approximately 20 gas turbine systems collected by Traverso[6] and correlated against TIT by ourselves. Values for each of the TITs used here appear in Table 3. This is only a crude way to specify blade cooling flow. More accurate blade coolant flow models can be found in Refs. 7 and 8. In addition to the major parameters discussed above, Table 4 lists some of the minor parameters that were used in the simulations.

Table 3 Blade coolant flow

Temp (°F)	Temp (°C)	Blade cooling flow (% of main input)
1,600	871	7.6198
1,700	927	8.2481
1,800	982	8.8871
1,900	1,038	9.5364
2,000	1,093	10.1955
2,100	1,149	10.8641
2,200	1,204	11.5420
2,300	1,260	12.2286
2,400	1,316	12.9239
2,500	1,371	13.6275

Table 4 Design parameters for turbine systems

Parameter	Value
Air intake	
Pressure	14.7 psia
Temperature	59°F
Relative humidity	60%
Compressor	
Isentropic efficiency	90%
Mechanical efficiency	99.8%
Inlet pressure drop	1%
Outlet pressure drop	0%
Combustor	
Pressure drop	4%
Heat loss	0.3%
Ratio of inlet fluid pressure to inlet combustor pressure	1.4
Expander	
Polytropic efficiency	85%
Mechanical efficiency	99.8%
Outlet pressure drop	0.5%

Compressor Surge

In a commercial turbine system, the compressors and expander are each designed for a particular combination of gas flow. All the water-augmented cycles affect this balance because they all add additional water to one or more of the gas streams.

This can lead to unstable operation of the compressor, called surge, which can result in catastrophic failure of the turbine system. Thus, for all the simulations of advanced cycles, an appropriate compressor has been matched to the existing expander.

THE CONVENTIONAL BRAYTON CYCLE

The Basic Cycle

Conventional cycles are Brayton cycles (Fig. 1). They consist of four major pieces of equipment: a compressor that compresses air to a high pressure; a combustor in which fuel is added and combustion takes place, producing a high-temperature gas; an expander, in which the gas is expanded to produce mechanical power; and a generator in which the mechanical power is converted to electric power. Some of the power produced in the expander powers the compressor. What remains above that is called the net power.

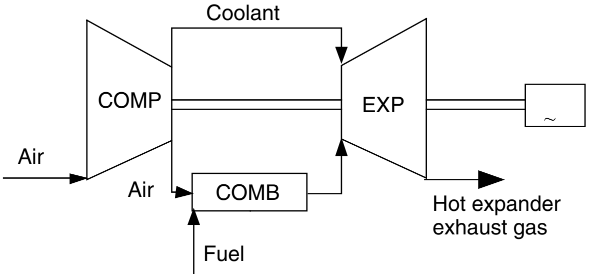


Fig. 1 The Brayton cycle.

The advantages of this cycle are:

1. It is simple. There is no additional equipment required beyond the four basic units.
2. It is proven technology. Most of the gas turbine systems in the world are of this design.

The disadvantages are:

1. The adiabatic flame temperature of a stoichiometric mixture of fuel and air is far too high for the expander blades to withstand. Thus, the fuel is mixed with a rather large amount of excess air in the combustion chamber to reduce the TIT to a value that is consistent with the requirements of expander blade integrity (1600°F–2500°F.) Because today's compressors and expanders are quite efficient (∼90%), this does not decrease the efficiency of the system greatly. Much of the power required to compress the excess air is recovered in the expander. But because this additional air has to be compressed and expanded, this does add considerably to the capital cost of the turbine system. Furthermore, even with excess air, at the high TITs of modern turbines, a certain amount of the gas produced by the compressor has to be used to cool the hot turbine blades. This results in something of a loss in turbine performance.
2. The exhaust gas exits at a high temperature; thus, a large amount of energy is wasted.
3. Pollutant gases, mainly CO and NO_x, are produced at unacceptably high levels. Additional technology needs to be used to control their levels.
4. The temperature of the gas leaving the combustor is difficult to control accurately. This causes operational problems in the control of the overall system.
5. At off-peak loads, the temperature of the gas entering the turbine drops. This cyclic temperature variation tends to increase corrosion and maintenance.
6. At off-peak electrical loads, the efficiency of the cycle also falls off quite rapidly. Because power generation systems often operate at less than peak load, this has an adverse effect on the average cost of the electricity produced.

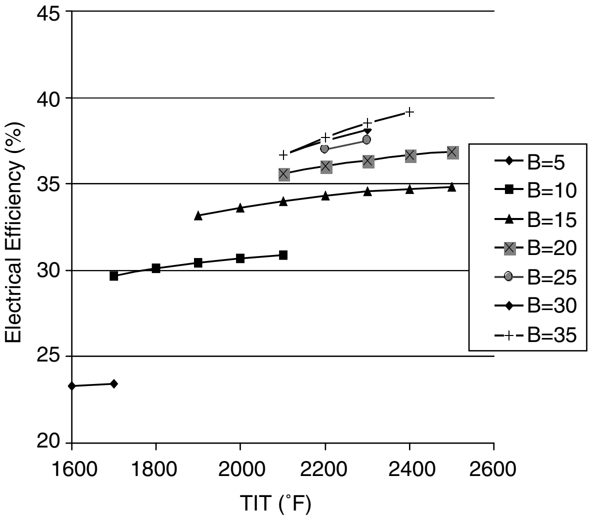


Fig. 2 Efficiencies of Brayton cycles.

The results of simulations for Brayton cycles for all the machines in Table 2 are shown in Figs. 2 and 3. Efficiency is not a strong function of TIT, but it rises as the pressure ratio increases. Specific power is a strong function of TIT, but it actually falls with the pressure ratio. The higher the TIT, the less excess air is needed for cooling, and this drives the specific power up. For larger machines, the best designs are at higher TITs and pressure ratios. For smaller machines, the capital costs are such that lower TITs and pressure ratios are more desirable.

Use of Recuperators

Whenever the expander exhaust temperature is higher than the compressor exhaust temperature, there is an opportunity to recover some of the exhaust heat by heat exchange with the compressed air before it enters the combustor (see Fig. 4). This has the effect of recycling some of the energy in the exhaust stream to enable it to be converted into mechanical energy. The cycles with pressure ratios of 20 and below were eligible for recuperation. In the simulations shown, it was assumed that the recuperators had 90% effectiveness. The effect of their inclusion on efficiency cycles at lower pressure ratios is dramatic (compare Figs. 5 and 2). There is a modest increase in air flow because at a higher temperature it has less cooling effect per unit mass. Thus, specific power drops somewhat (compare Figs. 6 and 3).

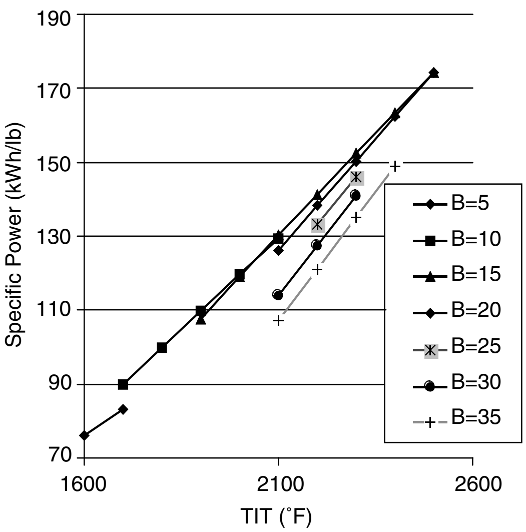


Fig. 3 Specific power of Brayton cycles.

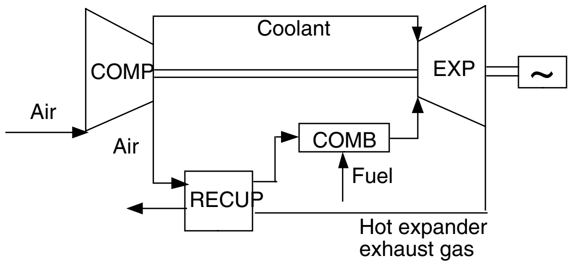


Fig. 4 Recuperated Brayton cycle.

Recuperators are gas-to-gas heat exchangers; therefore, they require very large surface areas, and are expensive and difficult to maintain. Where they can have great impact on efficiency—for small machines at lower pressure ratios—they comprise a large fraction of these machines' capital cost. Nevertheless, wherever recuperators can be used in power cycles, and to the extent that they can be made cost effective and trouble free, they represent a good way to improve cycle efficiency. This conclusion

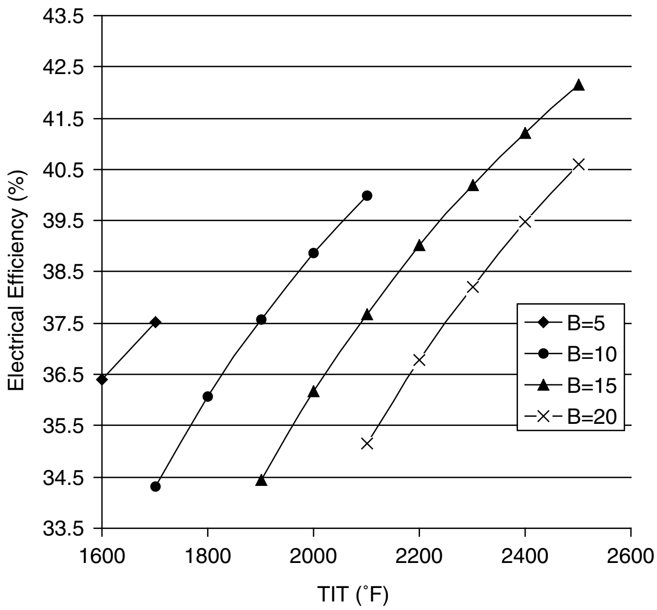


Fig. 5 Efficiencies of recuperated Brayton cycles.

Was–Wat

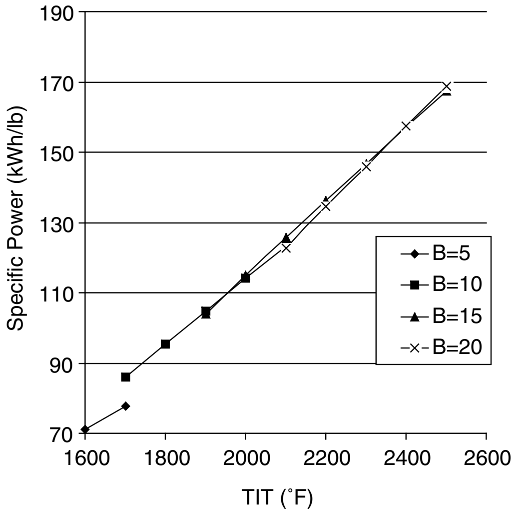


Fig. 6 Specific power of recuperated Brayton cycles.

applies not only to Brayton cycles, but also to most water-augmented cycles.

COMBINED CYCLES

The Basic Idea

The idea behind combined cycles (CCs) is to recover as much of the heat exhausted from a Brayton cycle (called the topping cycle) and use it to produce steam that is used to drive a steam turbine (the bottoming cycle). Fig. 7 shows a simple single-pressure CC. Heat is exchanged between the gas turbine exhaust and the water and steam of the steam turbine through a heat recovery steam generator (HRSG). This consists of three major units—an economizer to heat the colder water to near its boiling tempertuare, an evaporator to evaporate the water, and a superheater to heat the resulting steam to a higher

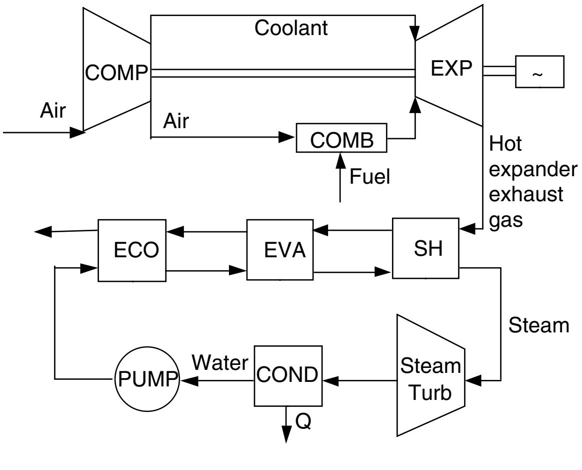


Fig. 7 Single-pressure combined cycle.

temperature. In actual HRSGs, there is also a deaeration step that takes place before the water enters the evaporator, but in these conceptual simulations, this was omitted. This does not affect the thermodynamic results significantly.

On the water/steam side, cool water is heated, evaporated, and superheated and then sent to the steam turbine, where the steam is expanded to subatmospheric pressure to produce mechanical energy. The steam is then condensed, pumped up to the high pressure, and reheated.

Heat Recovery Steam Generators (HRSGs)

Heat recovery steam generators are widely used to recover heat from hot gas streams. They have an inherent inefficiency in that evaporation takes place at a constant temperature, although it is driven by the hot exhaust gas, whose entry temperature is far above the boiling temperature of the water.

Heat transfer is limited by the so-called evaporator pinch temperature, which is the difference between the gas stream leaving the evaporator and the water's boiling temperature. Water is usually fed to the evaporator at a temperature slightly below boiling to prevent premature flashing. This difference is called the approach temperature of the liquid.

Evaluation of Combined Cycles

Combined cycles (CCS) are discussed extensively in a book by Kehlhofer et al.[9] This study used the midrange of his recommended approach temperature (8°F) and evaporator pinch temperature (12°F). It also uses a superheater pinch temperature of 18°F. These temperature differences are consistently used, wherever possible, in all HRSG designs for all of the cycles.

The advantages of CCs are:

1. There is a significant increase in efficiency (~15%).
2. Conventional gas turbine and steam turbine equipment can be used. The two cycles are somewhat independent, linked only by the HRSG, whose technology is also well known.

The disadvantages are:

1. There is a high capital cost because of the addition of a steam turbine. For this reason, such cycles have generally been implemented for larger power plants.
2. The HRSG is only moderately efficient in recovering gas turbine exhaust heat. The exhaust gas still exits at a relatively high temperature.
3. Steam turbine efficiency drops off at smaller sizes, which compromises the overall efficiency of this cycle at these sizes.

4. Disadvantages 1 and 3–6 of the Brayton cycle are unchanged.

The three base machines were simulated using a simple single-pressure steam bottoming cycle operating between two pressures suggested by Kehlhofer,[10] namely, 0.045 bar (=0.6586 psia) and 100.8 bar (=1475.3 psia). Summary results are shown in Table 5. In this table, the three base machines are identified by hyphenated numbers referring to the three most important characteristics of them three base machines, namely, base power, pressure ratio, and TIT. Because of its high efficiency, this is a widely implemented cycle. A number of variations have been developed that operate at up to three different pressures for even greater efficiencies. The high capital cost of this cycle has favored implementation in large power plants. However, the need for two separate turbine systems has motivated the search for cycles that do not require a second turbine but improve on the efficiency of the Brayton cycle. All the subsequent cycles considered here are examples of these.

WATER INJECTION INTO THE COMPRESSOR TRAIN

Quasi-Isothermal Compression

Isothermal compression up to the same pressure is known to require much less work than adiabatic compression to the same pressure. For example, for a compression ratio of 20 and $\gamma=1.4$, which is its value for air, and υ (isentropic efficiency)$=0.9$ the adiabatic $W/RT_0=5.56$, whereas the isothermal value is 3.33. (W/RT_0 is the non-dimensionalized adiabatic work.) Unfortunately, true isothermal compression is not possible to implement; therefore, two approaches, called quasi-isothermal compression, are often used. The first approach cools the gas between compression stages with conventional heat exchangers. The second approach injects water between stages. This is essentially a method of cooling because the absorption of the latent heat of evaporation of the water cools the gas.

Water-Injected Compression

Water injection into compressors is now being widely used both to improve cycle efficiency and to reduce emissions. Some early patents were obtained by Dow Chemical.[11] This idea has been incorporated into the general electric (GE) Sprint system.[12] Because water injection reduces the temperature of the compressor exhaust air, it can be combined with recuperation to produce efficient cycles. This is the idea behind the recuperated water injection (RWI) cycle developed by Rolls-Royce[13] (Fig. 8) and evaluated by Bassily.[14]

The Dutch utility company N. V. Kema took this idea one step further and obtained patents for what it calls the TOPHAT (Top Humidified Air Turbine) cycle with Swirlflash technology, which injects enough water to supersaturate the compressor air stream, thus ensuring a supply of water that evaporates continuously as the air passes through the compressor.[15]

The advantages of the RWI cycle are:

1. Significant gains in efficiency are achieved.
2. Relatively minor changes to the turbine system are necessary.

Its disadvantages are:

1. The amount of water that can be injected is limited by the amount required to saturate the air stream. Thus, the cooling effect is limited.
2. The use of a recuperator to recover some of the exhaust gas waste heat means that the temperature of the air that enters the combustor is higher than for a Brayton cycle, which requires somewhat more excess air for cooling.
3. The amount of waste heat recovered is limited by the temperature of the expanded air.
4. The water injected is lost as water vapor in the exhaust gas. This requires a continual use of treated water to be added into the cycle. There is usually not enough water injected to make it economical to recover it.
5. Although emissions are reduced through this technique, they are probably not reduced sufficiently to meet forthcoming standards.
6. Disadvantages 4, 5, and 6 of the Brayton cycle still apply.

Nevertheless, on balance and considering its relative simplicity, this is a remarkably efficient cycle, and it is surprising that it is not more widely implemented. In the simulations, it was found that water injection before the first compressor did not improve efficiency greatly; neither did the temperature of the injected water have much effect. The latent heat of evaporation dominated. In the simulated cycles, water was heated to 8°F below its boiling point before injection. It was assumed that the injected water led to 95% saturation of the air stream. Summary results are shown in Table 5.

HUMIDIFICATION OF AIR

The third major category of water-augmented cycles has as its main idea the very efficient humidification of the compressed air entering the combustor. The water used is pumped up to the system pressure and therefore does not need to be compressed as a gas. It replaces some of the

Table 5 Summary of performance for various cycles

	Brayton	Brayton w/ Recup.	CC	RWI	STIG-1:1	STIG-2.8:1	FSTIG	VAST	VASTIG
Efficiency									
46-10-19	30.46%	37.57%	44.96%	42.80%	31.44%	33.48%	40.48%	28.01%	35.26%
255-20-22	36.45%	37.21%	49.10%	46.94%	37.74%	40.32%	46.97%	36.46%	43.59%
1910-30-23	38.96%	N/A	49.55%	48.31%	40.37%	43.19%	48.85%	38.07%	46.49%
Power (kW)									
46-10-19	4,600	4,466	6,788	5,329	4,751	5,022	5,896	6,841	7,333
255-20-22	25,500	25,117	34,344	33,422	26,596	28,540	33,455	39,137	41,451
1910-30-23	191,000	N/A	242,953	276,734	200,565	217,785	252,696	305,975	326,852
Specific power (kJ/lb)									
46-10-19	109.8	104.8	146.5	132.0	114.5	123.9	157.7	205.3	228.8
255-20-22	139.9	136.1	172.3	196.6	147.6	163.0	206.4	268.8	295.7
1910-30-23	143.8	N/A	168.8	229.1	152.8	171.0	211.4	289.8	320.8
Water/fuel ratio									
46-10-19			6.63	5.82	1.00	2.80	8.97	8.94	12.74
255-20-22			5.90	6.11	1.00	2.80	7.40	7.52	10.85
1910-30-23			5.15	6.48	1.00	2.80	6.39	7.35	10.15

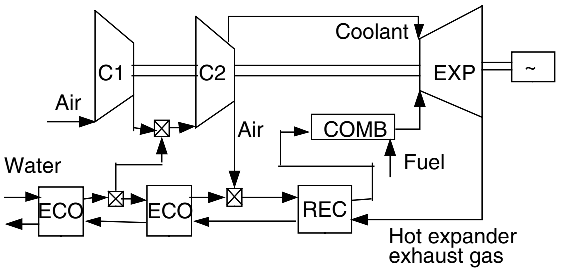


Fig. 8 RWI cycle (pumps not shown for clarity).

excess air and provides extra mass to be expanded. Furthermore, the humidification of the air—done relatively reversibly in a column—involves evaporation of water, which has a cooling effect on this water; therefore, the water can be used for interstage and aftercooling of the gas stream with surface heat exchangers.

This cycle has the most complex configuration of all the cycles considered (Fig. 9). The complexity comes from the pains that are taken to achieve thermodynamic efficiency. The distinctive feature of cycles in this family is the presence of a humidification column in which hot water flows downward, countercurrently, against rising cool air. Thus, the air is warmed and humidified. The configuration shown is a slight modification of the original humidified air turbine (HAT) cycle proposed by Ashok Rao,[16] which was an improvement over several previous schemes.

The entering air stream is both inter-and aftercooled with surface heat exchangers. The intercooling helps reduce the work of compression. It uses both recycled water, fresh feed water, and a heat sink to reduce these temperatures as much as possible. The aftercooling brings

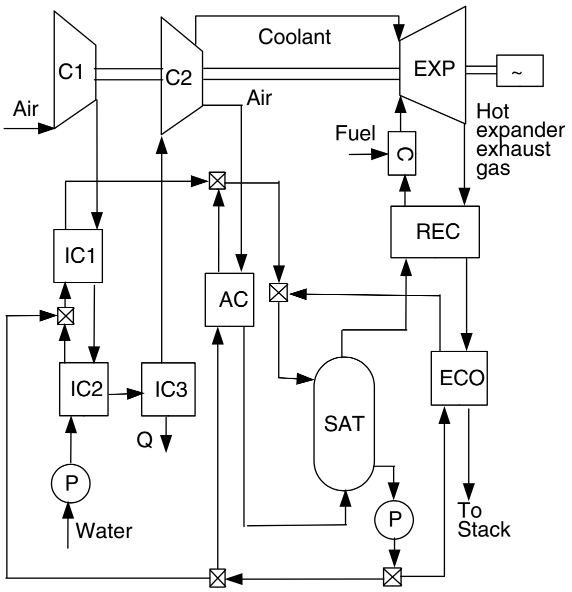


Fig. 9 HAT cycle.

the temperature of the air down to condition it for the humidification column. Hot water is passed countercurrently to the air stream and humidifies it. The hot water also warms the air, increasing its capacity to contain water vapor. The cooled water is used for cooling of the air stream, and it is also used in an economizer to capture and recycle expander gas heat. After humidification, the air stream passes through a recuperator and then to the combustor and expander.

A Swedish consortium consisting of Alstom's small turbine group, Lund Institute of Technology, the Royal Institute of Technology, and other educational and industrial organizations did research on this and related cycles.[17] They have proposed several modifications, which they called the evaporative gas turbine (EvGT) cycles.[18] Several additional studies were done by consortiums of companies and the Electric Power Research Institute (EPRI).[19,20] More recently, Hitachi has been building a non-intercooled demonstration plant (2–3 MW) in Japan.[21]

The efficiency of the cycle depends a great deal on the match between TIT and pressure, and also on the choice of temperatures and flow rates of the many streams in the cycle. Because of that, this entry doesn't report particular numbers, but Rao claims that in some configurations 65% efficiency is achievable, and our own calculations bear this out.

The advantages of this cycle are:

1. It has a very high efficiency and is also quite cost effective.
2. The water vapor in the air stream reduces NO_x emissions.
3. It is claimed that at off-peak operation, the TIT can be maintained somewhat.
4. It is claimed that at off-peak electrical load, the efficiency of the cycle stays essentially constant down to 60% of full load and decreases only to 74% of full load efficiency at 20% of full load, whereas in a CC, it can decrease to 59%.[22]
5. Efficiency and power output aren't greatly sensitive to changes in ambient temperature.
6. The amount of makeup water required is similar to what is required for a CC that uses a wet cooling tower.
7. The quality of the makeup water need not be extremely high, because it is not injected into the turbomachinery but evaporated in a saturator column.

The disadvantages are:

1. It is a complex cycle that seems to have deterred implementation.
2. Although the humidification column is similar to devices used in the chemical processing industry,

it is not as familiar in the power industry. The column also increases the capital cost of the cycle.
3. Although more water can be injected than in RWI, the amount is still limited to that which saturates the air stream. Thus, excess air is still needed to control the TIT.
4. The cycle incorporates a recuperator, which is a troublesome piece of equipment.
5. The temperature of the gas leaving the combustor is difficult to control exactly.
6. Disadvantages 4 and 5 of the RWI cycle still apply.

Nevertheless, this is an excellent family of cycles—especially for power production less than about 50 MW. It appears that the cycles' complexity has been the main obstacle to their implementation.

INJECTION OF STEAM AND WATER INTO THE COMBUSTOR

STIG Cycle

The last family of cycles consists of those that inject steam or water into the combustion chamber directly or immediately thereafter. This steam/water is produced by recovering heat from the expander exhaust with an HRSG, similar to what is done in the CC. Therefore, it suffers from the same HRSG efficiency limitation as the CC. Steam injection began with GE (steam injected gas turbines—STIG)[23] and Dah Yu Cheng,[24] who is the principal of Cheng Power Systems. The Cheng cycle is essentially very similar to the STIG cycle.

International Power Technologies (IPT)[25] now holds his major patents, which are listed on the Web.[26] International Power Technologies reports 128 systems installed worldwide, mostly in Japan.

"General electric in the mid 1980s had done a detailed study for Pacific Gas and Electric (PG&E) on the STIG and in the intercooled steam injected gas turbines (ISTIG) utilizing a modified LM5000 gas turbine. The steam to fuel ratio for the ISTIG was as high as 3.76 lb/lb fuel (natural gas) while cooling the stack gas to as low as 336 F (i.e., steam generation/injection was maximized). The resulting efficiency of the ISTIG was 54% on an LHV (lower heating value) basis."[22]

The STIG system makes steam in a manner very similar to the CC but injects it directly into the combustion chamber. Fig. 10 incorporates this cycle, although the injection of liquid water shown there is not done in STIG cycles. Cheng implemented his machines using existing turbine systems, especially the Allison (Rolls-Royce) 501-KX, which had a wide surge margin. He does not inject enough steam to lead to surge. Originally, this involved a

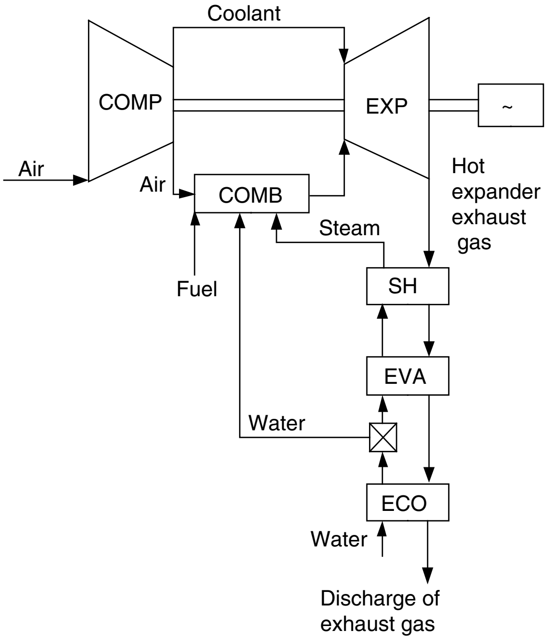


Fig. 10 VASTIG cycle.

steam/fuel ratio of 1:1, but 2.8:1 and higher is now state of the art. International Power Technologies claims that with the use of duct burners to increase the amount of steam that can be produced, it has achieved stable burning and no surge with steam/fuel ratios as high as 10:1.[27] Cheng has also developed a burner that he claims will reduce NO_x and CO emissions to below 2 parts per million (ppm). The Russian company Zorya–Mashproekt has developed what it calls the Aquarius cycle, which adds water recovery to a conventional STIG cycle.[28]

As with the CC, best results occur when steam is produced at the lowest possible pressure consistent with good injection into the combustor. Results from simulations of these systems are reported in Table 5.

The advantages of STIG cycles are:

1. The additional water mass is brought up to pressure as a liquid but is additional mass available for expansion.
2. Some of the excess air needed for cooling is replaced by water vapor.
3. Efficiency is improved.
4. Existing machines with a modified combustor can be used.
5. A burner has been redesigned, which not only permits combustion of the more humid gas, but also reduces emissions considerably.
6. Turbine inlet temperature can be maintained some degree at off-peak operation.
7. Off-peak efficiency is better than for the Brayton cycle.

Was–Wat

The disadvantages are:

1. Excess air is still required.
2. There is still heat lost in the exhaust gas.
3. The steam injected is usually lost as steam in the flue gas. This typically requires a continual use of treated water added into the cycle.
4. Control of the steam injection is a technical challenge.

FSTIG Cycle

In the STIG cycles, the superheater pinch temperature (18°F) can be reached, but the large pinch temperature in the evaporator suggests that theoretically, quite a bit more steam or water could be injected. The latter idea was explored by various investigators.[29,30] Urbach called the point of maximum steam injection the Cheng point. We have called this cycle full steam injected gas turbines (FSTIG), and performance results are also shown in Fig. 5.

The advantages of this cycle are:

1. Even more steam is introduced, thus accentuating the first three advantages of the STIG cycle.
2. There is now enough water in the exhaust gas to allow recovery and a water-neutral cycle, if that is desirable.
3. Turbine inlet temperature can be maintained further into off-peak operation.
4. Off-peak efficiency is better than for the STIG cycle.

However, there are some disadvantages (or at least problems) to be overcome:

1. The amount of steam injected will cause some existing turbomachinery to go into surge. Thus, a different compressor may have to be matched with the rest of the cycle or else other means have to be taken to prevent surge—but see the comments above concerning the STIG cycle.
2. The high concentration of water in the combustor exceeds the conventional combustion limit. However, VAST Power Systems (Value Added Steam Technologies) has developed a proprietary method of combustion that permits stable combustion at these high water levels and also reduces emissisons to the 5 ppm level.
3. Some excess air is still required.
4. The control problem is unchanged.

This cycle has not yet been commercially implemented, although it is a promising cycle because it brings steam injection to its limit and therefore improves efficiency.

The VAST Cycle

Water injection into the combustor goes back to 1950, where it was used for thrust augmentation for jet engines.[31] An early water-injection patent was obtained by Lyle Ginter,[32] and VAST Power Systems developed a cycle based on water-only injection into the combustor.[33] In Fig. 10, this corresponds to eliminating the steam injection and the superheater. The best results are obtained when the water is taken to the highest temperature before injection, which requires that it be brought to as high a pressure as is practical. Efficiency is low for such cycles because the water injection is a highly irreversible process. However, the significant reduction in capital costs makes such cycles competitive in cost effectiveness with other cycles of higher efficiency.

Its advantages are:

1. It is possible to eliminate virtually all excess air. This reduces the capital cost of the compressor and expander considerably.
2. Advantages 2–4 of the FSTIG cycle are maintained.
3. Water injection is easier to control than steam injection. Thus, the combustor temperature control point can be set closer to the desired TIT without as much concern for temperature overshoot.
4. With the addition of VAST combustion technology, stable burning can be achieved that reduces emissions to a very low level.
5. The injection of water tends to dampen acoustic vibrations.

The disadvantages are:

1. There is low efficiency.
2. Disadvantages 1 and 2 of the FSTIG cycle are retained.

The VASTIG Cycle

An obvious extension of the two previous cycles is to combine them as shown in Fig. 10. The name, VASTIG, indicates that this cycle is a combination of the VAST and STIG cycles. This idea was envisioned by Guillet[29] and Urbach.[30] As much steam as possible is injected, as in FSTIG. However, the exhaust gas still has heat that can be used to heat additional water in the economizer to produce enough water to replace virtually all the excess air. In many ways, this cycle combines the best of the two previous cycles. Results of the simulation of this cycle are shown in Table 5.

This cycle's advantages are the same as for the VAST cycle, with the additional advantage of higher efficiency

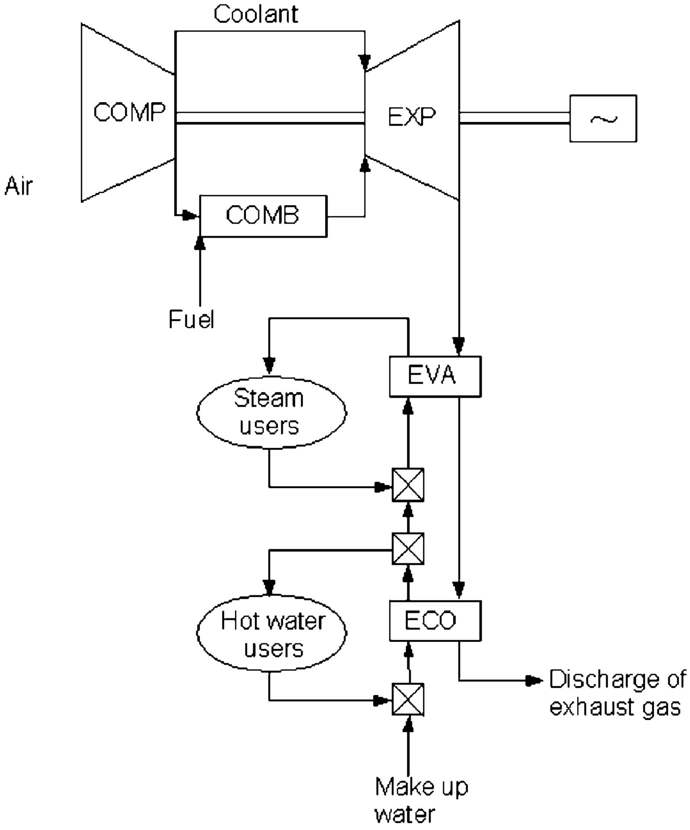


Fig. 11 CHP with Brayton cycle.

and better cost effectiveness. The disadvantages are the same as the first two for the FSTIG cycle.

It was thought that combining the VASTIG cycle with compressor water injection might lead to a very efficient cycle, but the efficiency of such a cycle was only marginally better than that of VASTIG alone.

COMBINED HEAT AND POWER

An idea that is coming into its own is that of producing both electric (or mechanical) power and heat (or refrigeration) with the same apparatus. This is called combined heat and power (CHP) or tri-generation when some of the energy is used for heating and some for chilling or refrigeration. Even the best gas-turbine power cycles recover only part of the chemical energy in the fuel. Some of the rest, which would otherwise be wasted, can be recovered as heat.

One straightforward and state-of-the-art method of doing this is illustrated in Fig. 11. The well-known Brayton cycle is equipped with an HRSG to recover heat from the exhaust gas in the form of steam or hot water. Table 6 shows how much heat can be recovered both in the form of steam and hot water. Such a cycle has a very high overall efficiency when one counts not only the electrical energy, but also the heat energy produced. A great deal of this energy is in the form of less desirable hot water rather than steam. However, there are refrigeration systems than

Table 6 Summary of performance of CHP cycles (pressure = 209.8 psia)

	Conventional Brayton	STIG 1:1	STIG 2.8:1
Electrical efficiency	30.36%	31.43%	33.46%
CHP efficiency	96.37%	95.88%	96.67%
Net power (kW)	4597	4749	5019
Steam heat (Btu/s)	6026	5142	3629
Steam temp. (F)	386	386	386
Steam heat/ total energy (%)	41.98%	35.91%	25.54%
Hot water heat (BTU/s)	3449	4086	5354
Hot water temp. (F)	262	190	171

(Continued)

Was–Wat

Table 6 (*Continued*)

Water heat/ total energy (%)	24.03%	28.54%	37.67%
Total heat (Btu/s)	9475	9228	8983
Total heat/ total energy (%)	66.01%	64.45%	63.21%

can be driven by hot water; therefore, this is a possible application.

Other cycles can be used for CHP with varying effectiveness. Table 6 also gives results for two STIG cycles. There are increases in electrical efficiency and power production. (Fig. 12)

CONCLUSIONS

Many cycles have been proposed that use water in imaginative ways to improve the efficiency and cost-effectiveness of power-generating cycles. Furthermore, some of these can be used with advantage in CHP systems as well. Cycles like these are only in the beginning stages of implementation, possibly facing the same kind of resistance that many new technologies face. However, world energy considerations and the need to reduce CO_2 emissions are likely to bring increasing pressure on utilities to generate power more efficiently. In the long run, more efficient systems will likely be based on one or more of the water-augmented systems that are just beginning to be implemented or are still on the drawing board.

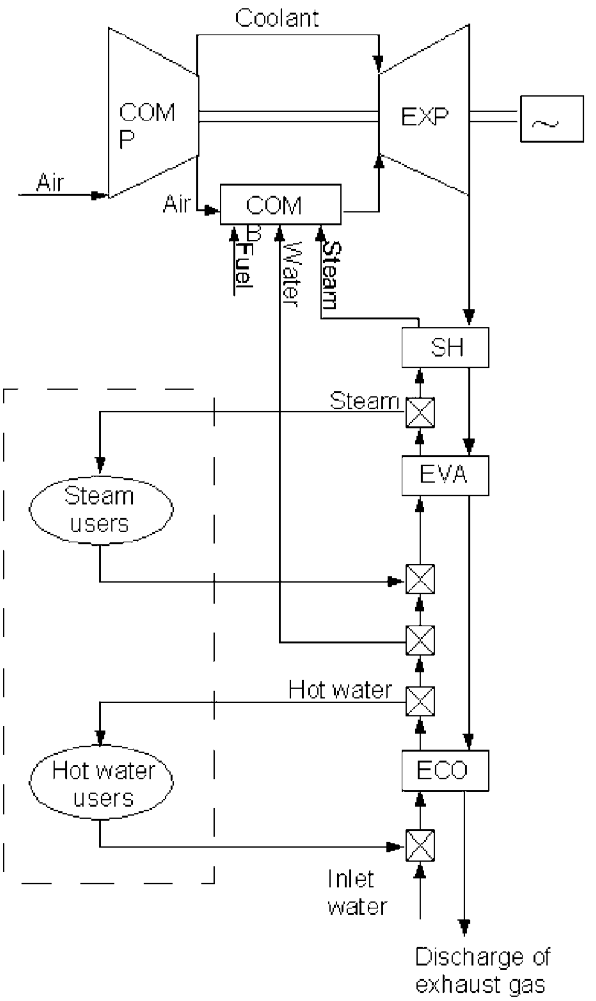


Fig. 12 CHP with VASTIG cycle.

ACKNOWLEDGMENTS

Helpful comments were made by Dr. David Hagen, VAST Power Systems, and Dr. Ashok Rao, University of California, Irvine. Allan McGuire assisted with some of the simulations. Support of this work by VAST Power Systems is gratefully acknowledged.

REFERENCES

1. Jonsson, M. *Advanced Power Cycles with Mixtures as the Working Fluid*, Doctoral thesis; Royal Institute of Technology, Dept. of Chemical Engineering and Technology, Division of Energy Processes: Stockholm, Sweden, 2003.
2. Klaus, R.L. Thermodynamics. In *Encyclopedia of Energy Engineering and Technology*; Capehart, B.L., Ed.; Taylor and Francis: NY, 2005.
3. http://www.technology.novem.nl/en/processtools/ (accessed July 2005).
4. http://www.simtechnology.com/IPSEpro/english/IPSEpro.php (accessed July 2005).
5. http://www.thermoflow.com/index.htm (accessed July 2005).
6. Traverso, A.; Massardo, A.F.; et al. *WIDGET-TEMP: A Novel Web-Based Approach for Thermoeconomic Analysis and Optimization of Conventional and Innovative Cycles*, ASME Paper GT2004-54115, Proceedings of ASME Turbo Expo 2004, Vienna, Austria, June 17, 2004.
7. Jordal, K. *Modeling and Performance of Gas Turbine Cycles with Various Means of Blade Cooling*, Doctoral thesis; Division of Thermoal Power Engineering, Dept. of Heat and Power Engineering, Lund Univ.: Sweden, 2001.
8. Torbidoni, L.; Horlock, J.H. *A New Method to Calculate the Coolant Requirements of a High Temerature Gas Turbine Blade*, Proceedings GT2004-53729, ASME Turbo Expo 2004, Vienna, Austria, June 14–17, 2004.
9. Kehlhofer, R.H.; Warner, J.; Nielsin, H.; Bachmann, R. *Combined-Cycle Gas Steam Turbine Power Plants*, 2nd Ed.; PennWell: Tulsa, OK, 1999.

10. Kehlhofer, R.H.; Warner, J.; Nielsin, H.; Bachmann, R. *Combined-Cycle Gas Steam Turbine Power Plants,* 2nd Ed.; Penn Well: Tulsa, OK, 1999; 51–53.
11. Zachary, R.E.; Hudson, R.D. Method and Apparatus for Achieving Power Augmentation in Gas Turbines Via Wet Compression. U.S. Patents 5,867,977, Feb. 9, 1999 and 5,930,990, Aug. 3, 1999, assigned to The Dow Chemical Company.
12. http://www.gepower.com/prod_serv/products/aero_turbines/en/lm6000_sprint.htm (accessed August 2005).
13. Reynolds, G.A. Power Generation Equipment. U.S. Patent 6,293,086, September 25, 2001, assigned to Rolls-Royce, London.
14. Bassily, A.M. Performance improvements of the recuperated gas turbine cycle using absorption inlet cooling and evaporative aftercooling. Proc. Inst. Mech. Eng., Part A: J. Power Energy **2002**, *216* (4), 295–306.
15. http://www.alphapowersystems.nl/Duesseldorf.pdf (accessed August 2005).
16. Rao, A. Process for Producing Power. U.S. Patent 4,829,763, May 16, 1989, assigned to Fluor Corp., now Fluor Daniels.
17. Lindquist, T. *Evaluation, Experience and Potential of Gas Turbine Based Cycles with Humidification*, Doctoral Thesis; Thermal Power Engineering, Lund Inst. Tech.: Sweden, 2002, http://www.lub.lu.se/cgi-bin/show_diss.pl/tec_560.html (accessed August 2005).
18. Bartlett, M. *Developing Humidified Gas Turbine Cycles*, Doctoral Thesis; Dept. of Chemical Engineering and Technology, Division of Energy Processes, Royal Institute of Technology: Stockholm, 2002.
19. Rao, A.D.; Francuz, V.J.; Shen, L.C.; West, E.W. *A Comparison of Humid Air Turbine (HAT) Cycle and Combined Cycle Power Plants*, EPRI Report EI-7300, March 1991.
20. Rao, A.D.; Francuz, V.J.; Mulato, F.J.; et al. *A Feasibility and Assessment Study for FT4000 Humid Air Turbine (HAT)*, EPRI Report TR-102156, Sept. 1993.
21. Hatamiya, S.; Araki, H.; Higuchi, S. *An Evaluation of Advanced Humid Air Turbine System with Water Recovery*, Paper GT2004-54031, Turbo 2004, Vienna.
22. Rao, A. Private communication.
23. http://www.gepower.com/prod_serv/products/aero_turbines/en/ (accessed August 2005); lm2500_stig.htm (accessed August 2005).
24. http://www.chengpower.com/ (accessed August 2005). See also Cheng, D.T.; Nelson, A.L.C. *The Chronological Development of the Cheng Cycle Steam Injected Gas Turbine During the past 25 Years*, Paper GT-2002-30119 at ASME Turbo Expo 2002, Amsterdam. A copy of this paper may be found at :http://www.aem.no/download/THE%20CHRONOLOGICAL%20DEVELOPMENT%20OF%20THE%20CHENG%20CYCLE%20STEAM%20.pdf (accessed August 2005).
25. http://www.intpower.com/ (accessed August 2005).
26. http://www.intpower.com/technology/patents.htm. (accessed August 2005).
27. Turley, R. *International Power Technology*, Private communication.
28. See http://www.mashproekt.nikolaev.ua/ and http://www.zorya.com.ua/Eng/Products/IndustrialEquipment/AquariusEng.htm (accessed August 2005).
29. Guillet, R. Natural Gas Stream Turbine System Operating with a Semi-Open Cycle. U.S. Patent 5,271,215, December 21, 1993.
30. Urbach, H.R.; Quandt, E.R. Direct Open Loop Rankine Engine System and Method of Operating Same. U.S. Patent 4.509,324, April 9, 1985.
31. Lundin, B.L. *Theoretical Analysis of Various Thrust-Augmentation Cycles for Turbojet Engines*, NASA Report TR-981; 1950.
32. Ginter, L. Engine System and Thermogenerator Therefore. U.S. Patent 3,651,641, March 28, 1972.
33. Ginter, L. Vapor-Air Steam Engine. U.S. Patents 5,617,719, April 8, 1997 and 5,743,080, April 28, 1998.

Water-Using Equipment: Commercial and Industrial

Kate McMordie Stoughton
Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.

Amy Solana
Pacific Northwest National Laboratory, Portland, Oregon, U.S.A.

Abstract
Water is an important aspect of many facets in energy engineering. The companion article of this entry, "Domestic Water-Using Equipment," detailed domestic related water-using equipment such as toilets and showerheads; this entry focuses on various types of water-using equipment in commercial and industrial facilities, including commercial dishwashers and clothes washers, single-pass cooling equipment, boilers and steam generators, cooling towers, and landscape irrigation. Opportunities for water and energy conservation are explained, including both technology retrofits and operation and maintenance changes. Water management planning and leak detection are also included because they are essential to a successful water management program.

INTRODUCTION

In the study of energy engineering, the importance of water management is often overlooked. But water is closely linked to energy. For commercial and industrial facilities, there are many pieces of equipment that utilize both water and energy to perform processes. For example, energy is required to circulate water through systems like cooling towers and to heat water for steam systems and boilers. In the development of a comprehensive energy management plan, it is important to be fluent in water management as well. Efficiency improvements in these types of systems typically work synergistically.

Common commercial and industrial water-using equipment are discussed below, with options for efficiency improvements and guidelines for estimating water consumption when not metered. In addition, leak detection and water management planning are discussed, and references are provided for further information.

DISHWASHERS

Dishwashing uses almost two-thirds of the water consumption in typical food preparation facilities such as restaurants.[1] Washing steps often include a scrapping trough system; a prewash; and a machine that cleans, rinses, and sanitizes. The scrapping trough system pushes garbage off the plates and into the disposer, using 3.0–5.0 gpm. The prewash removes anything left on the plates, using 3.0–6.0 gpm for both automatic and hand sprayer types. The final cleaning step uses 2.8–8.0 gpm.[2]

Retrofit options in the commercial setting are varied because of the multiple steps involved. The scrapping trough system could be used without water or replaced with a conveyor system, because food waste does not need to be disposed of in the sewer system. Prewash systems can be retrofitted with low-flow, high-pressure spray heads and flow reduction valves. A more efficient prewash is a manual high-pressure spray nozzle, which shuts off the flow when the user lets go and typically reduces water use to 1.8–2.5 gpm.[2] The most efficient models reduce flow to 1.6 gpm or less.[1]

A sensor could be installed on the dishwashing machine to ensure water flow only when dishes are passing through, instead of constantly or whenever the belt is in motion, empty or full. Low-temperature dishwashers are available that rely more on chemical agents instead of large volumes of hot water.[3]

Water recycling is another efficiency improvement. Rinse water could be reused for prewash or the garbage disposer. Dishwater could be reused in the garbage disposer. Water from final rinse steps could be reused in initial rinse or wash steps.

Auditing

To estimate the amount of water used in commercial dishwashing, information regarding flow rates of each component and frequency of use are needed. Flow rates of prewashers and sprayer nozzles may be measured as with faucets (see companion article, "Domestic Water-Using Equipment"), and information detailing the wash times and number of wash cycles of all equipment can be gathered. Use patterns need to be collected via interviews,

Keywords: Water; Dishwasher; Clothes washer; Single-pass cooling; Steam boiler; Cooling tower; Irrigation; Audit; Conservation.

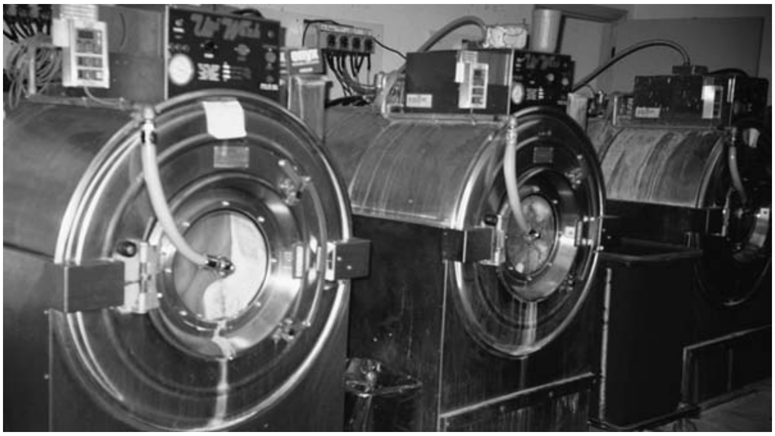


Fig. 1 Large commercial clothes washer.

including number of wash cycles per day and operating days per month.

CLOTHES WASHERS

Large commercial clothes washers (as shown in Fig. 1), on average, use between 2.5 and 3.5 gallons of water per dry pound of laundry.[2] A continuous batch (tunnel) washer reuses rinse water, which saves about 60% of the water typically used per load. Ozonated laundering systems use ozone to kill bacteria and disinfect fabric. They require no rinsing, hot water, or detergent, and are ideal for the sterilization of hospital gowns (for example), but are not suited for heavily soiled laundry (unless detergent is added).[4]

Large clothes-washing systems may be adapted to recycle rinse water for the next cycle's wash water, resulting in 25% or more water savings. Waste wash and rinse water can also be treated and then reused, allowing up to 50% water savings. Chemicals may be changed in the washing process to reduce the number of wash and rinse steps. Using only the necessary number of steps will help reduce water and energy use.[2] Find more information on laundry water recycling systems in the resource section of this article.

In any laundry facility or machine, washing only full loads will reduce overall energy and water use. Using cold water whenever possible saves energy by saving heated water.

Auditing

To determine laundry water use, machine frequency of use must be known. Industrial and some commercial laundry facilities usually operate on a regular schedule.

Once the number of loads per year is determined, the amount of water used per load can be estimated or found. Some washers have this value written on the machine. Others, usually in large commercial or industrial facilities, give the capacity in pounds of clothes. For machines with little or no information, manufacturer and model information can be used to obtain water use data. Incomplete data can be combined with typical consumption values based on the type of washer, as previously noted.

SINGLE-PASS COOLING EQUIPMENT

Single-pass cooling (also referred to as once-through or noncontact cooling) is a water-intensive cooling process that passes water once through equipment to remove heat before discharging it. Many different types of equipment may utilize single-pass cooling, including x-ray machines, hydraulic equipment, condensers, air compressors, welding machines, ice machines, air conditioners, and degreasers. These systems often run 24 h a day, 7 days a week, wasting thousands of gallons of water. A mere 2-gpm cooling rate would consume over 20,000 gallons per week alone.

To greatly increase efficiency, retrofit the single-pass cooling system to recirculate the water in a closed-loop system. If a closed-loop retrofit is not feasible, equipment should be operated at peak efficiency. This includes ensuring that the cooling process and water flow are turned off when not needed, accomplished either manually or through an automatic timer. Also, set the flow rate to the minimum allowed by the manufacturer's specifications if possible, and make sure the entering and leaving temperatures are in the specified range to ensure adequate cooling is taking place.

Routinely cleaning the heat exchange coils will maximize heat exchange, saving energy, and water. Consider using the effluent for cooling tower or boiler make-up water, landscape irrigation, or another nonpotable use. Another option is to replace the equipment with an air-cooled system to eliminate all water consumption.

Auditing

Information to gather that will help in the estimation of water and energy consumption includes the brand and model number, equipment time of use, entering and leaving temperatures, and flow rate. With this information, the manufacturer typically can provide the general range of cooling-water flow rates and energy requirements. The bucket-and-stopwatch method can also be used to estimate the flow rate of the cooling water (as explained in the faucet section of the companion article, "Domestic Water-Using Equipment"). Water consumption can be estimated by multiplying the operating hours by the flow rate. Energy consumption can be estimated by multiplying the gallons of water consumed by the energy required to cool 1 gallon to the working temperature, and dividing by the cooling equipment efficiency.

BOILERS AND STEAM SYSTEMS

Boilers and steam generators are addressed elsewhere in this book, but are discussed here because of their significant water use and conservation opportunities. These systems are typically used for process steam requirements, large heating systems, and cooking. Because the water is brought to such a high temperature in these systems, energy loss can be substantial in poorly operated equipment.

The largest water and energy losses are seen in systems without condensate return. In these systems, the condensed steam is sent down the drain after use instead of being recirculated. A continuous flow of fresh makeup water is required, and the system does not take advantage of any residual heat from the condensed steam, resulting in significant water and energy waste—up to 70% of operating costs.[2] Existing boilers and generators can be modified to return steam condensate if they do not already do so. Additional energy can be saved by insulating the condensate return pipes.

Systems with condensate return can still achieve efficiency improvements through proper operation and maintenance (O&M) of steam traps, lines, and blowdown. Steam traps can become stuck open, allowing steam to escape continuously instead of only when necessary as pressure builds. A maintenance program that includes regular steam trap checks and quick repairs helps to reduce these losses. Line leaks can also be moderated in this way.

Blowdown is necessary to control scale buildup inside the system. Phosphate-based chemicals and non-chemical methods are often used to reduce scale growth, but they cannot prevent it. Therefore, some water must be expelled and replaced periodically with make-up water. An automatic blowdown control manages this process to optimize water use, water quality, and boiler conditions.

Tracking the amount of make-up water a boiler consumes can help identify leaks in the system. This is true for systems with or without condensate return. Water meter data can show abnormal increases or decreases in consumption, indicating potential leaks. An increase may signify a leak in steam or condensate lines, while a decrease may signify a leak in a shell-and-tube heat exchanger. The latter type of leak is more problematic, because it means untreated water is entering the system, which can cause a boiler to fail.

Properly sized systems also can reduce water requirements, as can shutting down the system when not in use or during unoccupied hours. System shutdown can be manual or controlled automatically. Simply using the minimum amount of steam or heat required can greatly reduce the amount of water and energy used.

Auditing

Determining the amount of water used by a steam system can be difficult because it requires detailed system data; however, if a water meter is present on the make-up water line, the metered data can provide the consumption instead. Water and energy consumption can be calculated using some or all of the following information, depending on what is available: system capacity and time of use, steam pressures and temperatures, and fuel type and consumption. The operator and operating logs are the best source for most of this data, as well as manufacturer information found on the system and in manuals. Systems with condensate return use different calculations than systems without. Water lost through steam traps, leaks, and blowdown also needs to be identified and included.

COOLING TOWERS

Cooling towers (as shown in Fig. 2) are also discussed elsewhere in this book, but are briefly addressed here because of their considerable impact on water consumption. Because they provide cooling by evaporating water, conservation opportunities are minimal; however, some key O&M techniques can improve efficiency.

Cooling towers evaporate about 3 gpm through evaporation, drift, and blowdown (sometimes called bleed-off) for every 100 tons of cooling, with some variation by climate.[5] Reducing blowdown is the primary way to save water. Additionally, other uses may be found for blowdown water, and reclaimed water from single-pass cooling, reverse-osmosis systems, or other facility processes can be used as make-up water.

A high concentration ratio minimizes blowdown. The concentration ratio, or cycles of concentration, indicates how many times water circulates before being discharged. It is the ratio of the concentration of dissolved solids in the

Fig. 2 Cooling tower.

bleed-off water to the concentration of dissolved solids in the make-up water. The concentration ratio can also be calculated with the volume of the make-up and bleed-off water, if metered.

To increase the concentration ratio and save water, maintain water quality through chemical or other treatment methods:

- Shade cooling towers from sunlight by installing covers, which can significantly reduce biological growth, such as algae.
- Ultraviolet light added through an intense UV lighting module kills microorganisms that can build up and lead to fouling or even Legionnaire's disease.
- Sulfuric (or other) acid treatment lowers the pH of the water, which controls scale buildup created from mineral deposits.
- Sidestream filtration filters a portion of the flow to remove sediment and other impurities.
- Ozonation is a powerful oxidizer that controls scale, corrosion, and biological growth.
- Automatic chemical feed enables larger systems to automatically add chemicals to the system based on the conductivity, which optimally controls corrosion, scale, and biological growth.

Auditing

Metering is the best option to determine cooling tower water consumption. Meters that display total amount of water being used as well as current rate of flow are most useful. Measuring the water conductivity with a conductivity sensor is another way to understand the system. Conductivity measures how well electricity travels through the water, which is a function of the total ionic dissolved solids in the water. The ratio of make-up water conductivity to bleed-off water conductivity should be about the same as the ratio of make-up to bleed-off flow.[5]

LANDSCAPE IRRIGATION

Landscape irrigation can be a significant water consumer at many types of facilities. The amount used depends on the climate, water rates, type of landscape plants and plant health, type of irrigation equipment, and O&M practices. If nonpotable water cannot be used for irrigation, efficiency is essential.

The place to start improving efficiency is the landscape itself. Native turf and plants (an example of native southwestern landscape is shown in Fig. 3) have water and soil requirements that match the local average rainfall and soil type. Using these types of plants can greatly reduce or eliminate irrigation needs.

Healthy plants allow more efficient irrigation because the roots tend to be stronger and the plants tend to be drought and disease resistant. Frequent, shallow watering can result in compacted soil, shallow root systems, and increased disease incidence. Deep, occasional watering promotes deep, healthy roots. For turf, mulching mowers and high

Fig. 3 Native landscape in a dry climate.

blade settings leave a layer of protection to shade root systems, encourage deeper roots, and reduce evaporation.

Soil condition also plays an important role in plant health. Depending on the soil composition and plant type, organic matter can be added to the soil to help retain moisture and to ensure proper nutrients for the plant. Aeration loosens compacted soil, allowing more air and water to reach the roots; ideally, it should be performed at least twice per growing season (as shown in Fig. 4).

Optimized irrigation provides the landscape with the amount of water it actually needs. The evapotranspiration (ET) rate, (ET, the loss of water by evaporation and transpiration, indicates how much water is needed by the plant for the current conditions based on plant type, soil type, and season. Local county extension offices or local university agricultural departments often have information on how to estimate ET rates.) typically provided in inches of water, is compared to the area's rainfall and soil type to

Fig. 4 Turf aeration.

find the amount of irrigation needed. The soil type determines how much moisture can be held at the plant's root base (sandy soil will retain much less water than clay soil) and what types of nutrients might be deficient. Divide the ET by the irrigation equipment's precipitation rate (PR), or flow rate, to calculate the watering time needed. Precipitation rate is typically provided in inches per hour, and can be obtained from the manufacturer.

Appropriate irrigation scheduling can then provide the landscape with the required amount of water. Irrigation scheduled in the early morning helps to reduce evaporation and minimize fungus growth. This time period is typically less windy as well, helping to avoid overspray and evaporation. Overspray to sidewalks and pavement can also be avoided by placing sprinkler heads only where they will evenly distribute water over the desired area.

Irrigation controllers function automatically and allow flexibility in scheduling. Automatic timers are helpful for watering during inconvenient times of day, and for preventing overwatering because of negligence of sprinkler shutoff. A manual override option is important for factors where irrigation is unnecessary, such as rainfall.

Other technologies, including rain, wind, freeze, and humidity sensors, respond to weather conditions and can be connected to the existing control system. An ET controller is a more robust option, which can be programmed to match the plant's exact water needs with the PR of the irrigation system and the local weather conditions.

Efficient irrigation equipment applies water directly to the plants' root base and avoids overspray and evaporation. Low-volume drip systems apply water at a low pressure to nonturf landscape such as flowerbeds and shrubs. Subsurface drip irrigation applies water directly to roots of turf or other plants through a perforated piping system buried underground.

Proper maintenance of irrigation equipment can also save water. System checks for broken heads and line leaks at the beginning and end of each watering season are crucial. Routine spot checks of the irrigation system are also important, especially if landscape is watered early in the morning when broken heads and leaks can go unnoticed. The system's pressure should meet the range of pressures specified by the manufacturer. If the pressure is too high, a pressure-reducing valve might be required, which also needs to be checked periodically for leaks.

Auditing

Metering is the best approach to measuring irrigation water consumption. This way, the consumption can be tracked and monitored to ensure the irrigation needs are as expected. If metered data are not available, system operators should be able to estimate how long the irrigation system runs each day, week, and year. If the system flow rate is unknown, it may be found in manufacturer data. Alternatively, consumption can be estimated using the following typical values:

- Water-thirsty landscape: 25 gallons per square foot per year.[2]
- Water-wise landscape: 5–10 gallons per square foot per year.[2]

LEAK DETECTION

Distribution system leaks can be a large source of water loss, especially in aging infrastructure. Leaks typically occur at joints and connections such as meter boxes, hydrants, and valves; resulting from loose connections, corrosion, improper piping installation, settlement, or overloading. Leak rates have a wide range; a large main break can lose water at 1000 gpm, while a slow service line leak can be as small as 0.5 gpm.[6] These small leaks will most likely grow over time, especially if they are caused by corrosion.

A comprehensive leak-detection program is the best approach to limiting system losses. A good program includes the following steps:

- Estimate the total loss rate through a distribution system audit.
- Identify the areas of highest leak rate.
- Inspect and record the location and condition of service lines, valves, meters, and other connections.
- Prepare a detection-and-repair plan that schedules leak inspections and repairs and analyzes economics.
- Implement the plan and revise when needed.

A simplified technique for finding leaks is called the zone-flow procedure, which examines water use in specific areas of the system.[6] Zones of the distribution system are individually isolated by closing the valves in that area, and flow is recorded for about 24 hour. If flow exists during the nighttime hours but there are no major nighttime water-using processes, leaks are most likely present in that zone.

Leaks are most often pinpointed by using electronic listening devices, which detect the sound of the pressurized water being forced through the pipe wall.[6] Typically, a trained operator is contracted to perform leak-detection services, because it requires specific instrumentation and knowledge of how to determine the exact location of the leak.

Thermal imaging can help find leaks in large distribution lines (16 in. or greater) using multispectral and thermal aerial photography. Areas of cooler temperatures reveal possible water leaks because the saturated area will be cooler than the surrounding soil. This technique works best during warmer seasons and in arid areas for two reasons: (1) the leak is easy to pinpoint because vegetation grows at the point of the leak; and (2) the water table is typically deeper, so the surrounding soil temperatures are warmer than the area of leakage.

WATER MANAGEMENT PLANNING

Water management planning is at the heart of all water efficiency improvements, and sets water conservation as a priority alongside energy efficiency. Ideally, a comprehensive water management plan uncovers the entire cycle of water use at a facility from supply through discharge (as shown in Fig. 5 below).

A tiered approach can help develop a comprehensive water management plan, as described in the following steps:

1. Assess current conditions:
   - Understand and apply water-related regulations and laws.
   - Work with stakeholders to understand issues revolving around different aspects of water management.
   - Gather comprehensive utility information, including rates and at least one year's bills. Include information on rate structure and future water rates.
   - Perform audits to quantify water consumed by equipment, other end-uses, and leaks or losses. Use auditing techniques (described in both water-related articles) to help estimate water consumption at the equipment level.
   - Document current O&M practices and schedules.
   - Categorize and quantify water consumption by end use to help prioritize efforts.
2. Improve operations and increase efficiency:
   - Identify potential operational efficiency improvements and technology retrofits for each major water-using equipment and process.
   - Perform economic analyses to identify feasible projects.
   - Prioritize efficiency improvements based on largest water user and best economic feasibility.
   - Set goals and a schedule for project completion.
   - Ascertain funding to implement projects.
   - Share success.
3. Plan for contingencies:
   - Assess current contingency plans for emergencies such as short term outages and drought.
   - Develop a proactive plan that prepares and develops strategies for contingencies. This includes assigning specific tasks to teams that are responsible for monitoring and mitigating water shortages and outages.

Understanding current consumption patterns is a crucial element to successful water management. Submetering buildings and end uses is the best and most accurate method to allocate water use. When submetering is not feasible, manual measurements or engineering estimates can be used as an alternative.

When deciding which conservation opportunities to pursue, a good resource is the U.S. Department of Energy Federal Energy Management Program's (FEMP) "Water Efficiency Best Management Practices" (BMPs), which are recommendations for efficiency improvements in ten

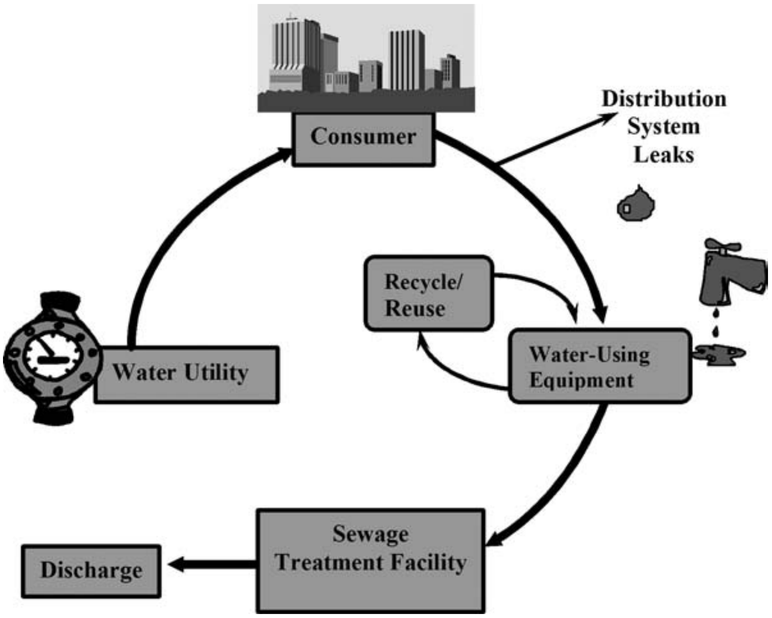


Fig. 5 Facility water cycle.

Was–Wat

distinct areas. In the "Resources" section below, a link is provided to the website.

CONCLUSION

Water-using technologies are an important aspect of energy engineering, because their use has such a wide, but often unrealized, impact on energy. Because unrestrained water consumption is becoming more of a concern, more technologies are being developed to help conserve both water and energy. Simple retrofits and proper O&M will make a significant impact on water consumption, especially in large water-consuming systems found in commercial and industrial facilities. These changes begin with organized water management planning and a comprehensive audit; conservation follows.

RESOURCES

General

American Water Works Association
www.awwa.org
Information on AWWA's activities, programs, conferences, publications, and useful links.

WATER EFFICIENCY AND CONSERVATION

Environmental Protection Agency Water Efficiency Program
www.epa.gov/owm/water-efficiency/index.htm
Efficiency tips on agriculture, residential water use, drought management, water recycling, and more.
California Urban Water Conservation Council
www.cuwcc.org
Statewide water conservation awareness program with urban water conservation best management practices.

Water and Energy

State of California Energy Commission Energy and Water Connection
www.energy.ca.gov/process/water/water_index.html
Information specifically tying together energy and water.
Watergy
watergy.org
A division of the Alliance to Save Energy, drawing the links between water supply and distribution and energy.

Federal Guidance

Federal Energy Management Program—Water Efficiency
www.eere.energy.gov/femp/technologies/water_efficiency.cfm
Guidance to federal agencies on water efficiency improvements.
Federal Water Efficiency Best Management Practices
http://www.eere.energy.gov/femp/technologies/water_fedrequire.cfm.
Listing of ten key areas for improving efficiency in water-using equipment and processes.
Energy Center of Expertise—Water Management Guide
www.gsa.gov/gsa/cm_attachments/GSA_DOCUMENT/waterguide_new_R2E-c-t-r_0Z5RDZ-i34K-pR.pdf
A comprehensive handbook for federal facility managers on water conservation.

Clothes Washers

http://www.h2oreuse.com/WhatHow.html
Information on a laundry water recycling system.

Cooling Towers

Federal Technology Alert: Ozone Treatment for Cooling Towers
http://www.eere.energy.gov/femp/pdfs/FTA_OTCT.pdf
Technical bulletin on ozonation for cooling towers.
Labs for the 21st Century Best Practices for Water
http://www.labs21century.gov/pdf/bp_water_508.pdf
Resources on water management in federal laboratories.

Irrigation and Landscape

Irrigation Association
www.irrigation.org
Irrigation best management practices and information on auditor training classes.
Center for Irrigation Technology
cati.csufresno.edu/cit/index.html
Independent research and testing of irrigation technology.
GreenCo
www.greenco.org
Information on creating drought-resistant landscapes through best management practices.
Sustainable Building Xeriscape Sourcebook
www.greenbuilder.com/sourcebook/xeriscape.html
Sustainable building sourcebook on xeriscape principals and construction specifications.

ACKNOWLEDGMENTS

The authors would like to thank Bill Chvala for his technical contributions, Mark Halverson for his technical peer review, and Sue Arey for her editorial review.

REFERENCES

1. http://www.cuwcc.org/sprayvalves.lasso (accessed on January 2006).
2. A Water Conservation Guide for Commercial, Institutional and Industrial Users, New Mexico Office of the State Engineer, Santa Fe, New Mexico, 1999; pp. 38–68.
3. Water Management Guide, U.S. General Services Administration, Kansas City, Missouri, 2000; pp. 66–67.
4. McMordie Stoughton, K.L.; Solana, A.E.; Elliott, D.B.; Sullivan, G.P.; Parker, G.B. In *Update of Market Assessment for Capturing Water Conservation Opportunities in the Federal Sector*, *PNNL-15320*; Pacific Northwest National Laboratory: Richland, Washington, 2005; p. E.4.
5. McMordie Stoughton, K.L.; Chvala, W.D.; Solana, A.E.; Halverson, M.A. In *Actions You Can Take Now to Save Water and Reduce Costs*, *PNNL-SA-45992*; Pacific Northwest National Laboratory: Richland, Washington, 2005; pp. 7–9.
6. Water Audits and Leak Detection, Vol. M36, 2nd Ed., American Water Works Association, Denver, Colorado, 1999; pp. 38–40.

Water-Using Equipment: Domestic

Amy Solana
Pacific Northwest National Laboratory, Portland, Oregon, U.S.A.

Kate McMordie Stoughton
Battelle Pacific Northwest National Laboratory, U.S. Department of Energy, Richland, Washington, U.S.A.

Abstract
Water management is an important aspect of energy engineering. This entry addresses water-using equipment used primarily for household purposes, including faucets, showers, toilets, urinals, dishwashers, and clothes washers, and focuses on how this equipment can be optimized to save both water and energy. Technology retrofits and operation and maintenance changes are the main water and energy conservation methods discussed. Auditing to determine current consumption rates is also described for each technology.

INTRODUCTION

In the study of energy engineering, the importance of water is often overlooked, although energy and water are closely linked. Water not only generates energy (as hydropower), but also consumes energy. The energy required to heat water used in sinks and showers, for instance, directly correlates to the amount of water being used. All potable water (water safe for human consumption) is treated and pumped before and after use, requiring energy for the pumping, monitoring, and production and transport of chemicals.[1]

Water is important regardless of its ties to energy. Droughts are having a greater impact on wildlife, animal and fish habitats, and humans; they can easily result in water shortages at the current level of consumption.[2] In addition, rising water rates are a major concern for consumers. Water conservation has become a necessity and should be interrelated with energy efficiency, helping to conserve natural resources.

Common domestic water-using equipment is discussed below, namely faucets, showerheads, toilets, urinals, dishwashers, and clothes washers. These technologies do not consume large amounts of water alone, but their combined daily use significantly impacts available resources. Water consumption can be greatly reduced by installing and using the most efficient appliances and water-using systems. Proper operation and maintenance (O&M) is even more essential to minimizing water and energy use. Options for efficiency improvements and guidelines to determine water consumption for non-metered domestic end uses are addressed in this section.

Keywords: Water; Faucet; Shower; Toilet; Urinal; Dishwasher; Clothes Washer; Audit; Conservation.

FAUCETS

A variety of faucets on the market today fall into two main categories: manually operated or automatic closure. As a result of the Energy Policy Act of 1992 (EPAct 1992), faucets are mandated to remain below a flow of 2.2 gallons per minute (gpm) at 60 pounds per square inch (psi); however, flow rates as low as 0.5 gpm are available. Older faucets flow at about 3–5 gpm at 60 psi. Automatic sensor faucets typically run at 0.5 gpm[3] and use infrared or ultrasonic sensors to turn on the faucet in the presence of a person. These sensors are powered by hard-wired, solar-powered, or hydro-powered electricity. The solar- and hydro-powered versions use batteries that are recharged with ambient light and running water, respectively. Self-closing faucets shut off the water when the knob is released, and typically have a 2.2-gpm flow rate. Metered valve faucets release a preset amount of water; the EPAct 1992 standard is 0.25 gallons per use.[4] Foot pedal faucets are retrofits to existing faucets that allow hands-free operation.

When replacing faucets with new fixtures, use the lowest flow possible for faucets that are not used for filling containers. If sanitation is a concern, consider automatic sensors or foot pedals.

Low- and no-cost options that do not replace the entire faucet also reduce water consumption. Aerators (see Fig. 1) are inexpensive and simple to install, and can reduce the flow of an existing faucet to 2.2 gpm, or as low as 1.0 gpm.[3] Flow restrictors are washer-like devices that fit into a faucet head. The flow valve on the water line can also be adjusted to reduce flow. Regular maintenance, which includes checking for leaks and proper line pressure (30–60 psi) and making appropriate adjustments, will save water. High pressures can waste water and quickly erode fixtures.

Fig. 1 Faucet aerator.

Auditing

Some faucets and aerators have flow rates imprinted on the spout. These numbers can be used if available, but are often not accurate. A simple measuring method is the bucket and stopwatch method, which measures the time that it takes to fill a container of known volume, allowing the gallons per minute to be calculated. A bag with volume increments marked on it can be filled for 5 seconds and the volume read to determine the gallons per minute; these supplies are available through water conservation suppliers.

The flow rates are then multiplied by the amount of time each faucet is used. Building occupancy (hours occupied and number of people, split by gender if possible) determines this. The following average values for faucet run times during an 8-hour work day can be used:

Men: 1.2 min/day
Women: 2.0 min/day[4]

Daily and annual water use per person for common faucet flow rates is summarized in Table 1. The table also shows the savings potential that can be achieved by retrofitting an older faucet with a more efficient one. This table is based on a 40-hour work week.

When determining energy savings associated with faucet water savings, it can be assumed that about 50% of the total flow is hot water.[6] When heating water from 60 (typical delivery temperature) to 120°F, electric hot water heaters use about 0.147 kilowatt-hours (kWh) per gallon, and natural gas heaters use about 0.005 therms per gallon. Energy savings can be estimated using the following equation with these assumptions:

$$\text{Energy saved} = \frac{\text{Water saved (gal)} \times \%\text{ hot water} \times \text{energy used (kWh or therms/gal)}}{\text{heater efficiency}}$$

Daily and annual energy use per person is summarized in Table 2, using the same faucet types as in Table 1. This table also shows the energy savings potential from the water savings calculated in Table 1.

SHOWERS

Energy Policy Act 1992 standards require a 2.5 gpm flow rate, but older showerheads are still widely utilized today, at an average of 5–7 gpm. Low-flow showerheads are available with flow rates as low as 2.5 gpm. These do not always perform well, resulting in a somewhat negative reputation. However, water-conserving showerheads offering strong performance are widely available.

The same water conservation options for faucets exist for showers. Flow restrictors are extremely cost-effective, but can result in poor water pressure.[5] Showerheads with temporary cutoff valves are another option (see Fig. 2); water can be shut off while soaping and turned on to rinse at the same temperature. Additionally, using individual control valves in common shower rooms saves water compared with a single controller, which turns all showers on or off at the same time.

Table 1 Estimated water consumption and savings by efficient faucets in commercial facilities

Fixture/ year of installation	Water use (gpm)	Daily consumption male/ female (gal/day)		Annual consumption male/ female (gal/year)		Annual savings—Energy Policy Act (EPAct) standard retrofit male/ female (gal/year)		Annual savings—low-flow retrofit male/female (gal/year)		Annual savings—high efficiency retrofit male/female (gal/year)	
High efficiency	0.5	0.6	1.0	156	260	—	—	—	—	—	—
Low-flow	1.0	1.2	2.0	312	520	—	—	—	—	156	260
EPAct standard	2.2	2.6	4.4	676	1,144	—	—	364	624	520	884
Pre-1994	3.0–5.0	3.6–6.0	6.0–10.0	936–1,560	1560–2,600	260–884	416–1,456	624–1,248	1,040–2,080	780–1,404	1,300–2,340

Table 2 Estimated energy consumption and savings by efficient faucets in commercial facilities

Fixture/ year of installation	Energy use (kWh/min)	Daily consumption male/ female (kWh/day)		Annual consumption male/ female (kWh/year)		Annual savings—Energy Policy Act (EPAct) standard retrofit male/female (kWh/ year)		Annual savings—low-flow retrofit male/female (kWh/ year)		Annual savings—high efficiency retrofit male/female (kWh/year)	
High efficiency	0.04	0.04	0.07	12	19	—	—	—	—	—	—
Low-flow	0.07	0.09	0.15	23	38	—	—	—	—	12	19
EPAct standard	0.16	0.19	0.32	50	84	—	—	27	46	38	65
Pre-1994	0.22–0.37	0.26–0.44	0.44–0.74	69–115	115–191	19–65	31–107	46–92	76–153	57–103	96–172

Fig. 2 Temporary cutoff valve showerhead.

Auditing

Similar to faucets, shower water consumption can be found using the bucket and stopwatch method, or by checking the manufacturer's rate imprinted on the showerhead. Building occupancy may or may not be representative of the number of showers taken daily. This should be confirmed during an interview of facility staff or occupants. Some occupants may not shower every day; some may shower elsewhere, like a health club or workplace.

The length of showers can vary depending on the facility and user. Five-minute showers can be used as a conservative estimate, but interviews may help determine a more accurate length. Daily and annual water consumption per person for common showerhead flow rates is summarized in Table 3. The table also shows the savings potential that can be achieved by retrofitting an older showerhead with a more efficient one. This table assumes a person showers once every day for 5 minutes.

About 60% of a typical shower flow rate is hot water.[6] Otherwise, energy savings is calculated in the same manner as for faucets. Table 4 summarizes daily and annual shower energy use and the energy savings potential based on the water savings in Table 3.

TOILETS

Toilets are the largest indoor water consumer at most facilities.[7] The most common types of toilets on the

Table 3 Estimated water consumption and savings by efficient showerheads

Fixture/year of installation	Water use (gpm)	Daily consumption (gal/day)	Annual consumption (gal/year)	Annual savings—Energy Policy Act (EPAct) standard retrofit (gal/year)	Annual savings—low-flow retrofit (gal/year)	Annual savings—high efficiency retrofit (gal/year)
High efficiency	1.5	7.5	2,738	—	—	—
Low-flow	2.0	10.0	3,650	—	—	912
EPAct standard	2.5	12.5	4,563	—	913	1,825
Pre-1994	5.0–7.0	25.0–35.0	9,125–12,775	4,562–8,212	5,475–9,125	6,387–10,037

Table 4 Estimated energy consumption and savings by efficient showerheads

Fixture/year of installation	Energy use (kWh/min)	Daily consumption (kWh/day)	Annual consumption (kWh/year)	Annual savings—Energy Policy Act (EPAct) standard retrofit (kWh/year)	Annual savings—low-flow retrofit (kWh/year)	Annual savings—high efficiency retrofit (kWh/year)
High efficiency	0.13	0.66	242	—	—	—
Low-flow	0.18	0.88	322	—	—	80
EPAct standard	0.22	1.10	403	—	81	161
Pre-1994	0.44–0.62	2.21–3.09	805–1,127	402–724	483–805	563–885

market are tank toilets (see Fig. 3), typical in residential and light commercial applications, and flush valve toilets (see Fig. 4), typical in commercial applications. Tank toilets flush water through a siphon created by emptying water into the bowl through a flapper valve, whereby water is sucked out of the bowl and down the sewer line. Flush valve (or flushometer) toilets use the line pressure to flush high velocity water through the toilet; line pressure of at least 30 psi is required.

Pressure-assisted toilets contain a chamber with trapped air inside the toilet tank. The trapped air compresses as the chamber fills with water. When the toilet flushes, the compressed air pushes the water out at a high velocity, providing a powerful flush. Dual flush toilets provide two flushing options: a full flush and a partial flush. These toilets are available as tank, pressure-assisted, or flush valve types.

Toilets manufactured between 1980 and 1994 typically consume about 3.5 gallons per flush (gpf); toilets manufactured prior to 1980 consume around 5.0 gpf. Energy Policy Act 1992 mandated that toilets do not exceed 1.6 gpf; these toilets were termed ultra low-flow toilets or ULFs. High efficiency toilets (HETs) have since emerged on the market, defined as consuming at least 20% less than standard 1.6 gpf toilets. High efficiency toilets on the market include 1.0 gpf pressure-assisted and flush valve toilets and dual flush toilets, which use 1.6 gpf for the full flush and between 0.8 and 1.1 gpf for the partial flush. The flushing performance of low-flow toilets has been a consumer concern in the past, but has since improved.

An option for increasing the efficiency of an existing flush valve toilet is a flush valve retrofit kit. This kit involves replacement of the existing valve and toilet bowl

Fig. 3 Inside of tank toilet.

Fig. 4 Flush valve toilet.

with a high efficiency valve (rather than the entire plumbing) and new bowl. Toilets will not have a proper flushing pattern if the bowl is not replaced with the valve retrofit.

Retrofit devices for water efficiency in tank toilets include displacement devices, toilet dams, dual flush adapters, and early closure devices. These devices either displace volume in the tank or change the flushing mechanism to minimize the amount of flush water. Water utilities commonly provide these types of devices as a low cost retrofit option, but flush performance and long term maintenance may be problematic.

Leaks are quite common in toilets and are often inaudible, allowing them to go undetected for months, possibly years. A good O&M program can go a long way to save water, usually at little or no cost. Some simple procedures to include are as follows.

- Periodically check for leaks, at least bi-annually.
- Replace worn valves and flapper valves, matching the proper flush rate for the existing toilet. For example, if a 1.6-gpf tank or flush valve toilet is retrofitted with a 3.5-gpf flapper valve, the toilet will flush with 3.5 gallons.
- Periodically sample flush rates.
- Avoid harsh chemicals dispensed through tank toilets to avoid degrading flapper materials prematurely.

Auditing

Toilet water consumption begins with vintage identification to find the consumption per flush described above. The total water use can be found by understanding the number of people using the toilets and the total number of uses by each person. Daily and annual water use for each toilet vintage is summarized in Table 5, based on common estimates of toilet use frequency for one man and one woman. The table also shows the savings potential that can be achieved by retrofitting an older toilet with a ULF or HET. This table is relevant for workplaces with a 40-hour work week; because men have access to urinals, it is assumed they use toilets once per day and women use toilets three times per day.

URINALS

Urinals are typically used in commercial applications only. There are two basic types of urinals: flush valve (in Fig. 5 with infrared sensor) and no-water. A flush valve urinal works the same way as a flush valve toilet (described above). The no-water urinal is distinctly different, as the name implies. Instead of flushing, it contains a cartridge in the drain with a sealing liquid less dense than urine, allowing the urine to pass into the drain line. The liquid then seals the cartridge and prevents sewer vapors from escaping back into the restroom.

Similar to toilets, the amount of water a urinal uses depends on its manufacture date. Urinals built prior to 1980 use approximately 5.0 gpf. Urinals manufactured between 1980 and 1994 consume 1.5–3.0 gpf. From 1994 forward, EPAct 1992 required urinals to flush no more than 1.0 gpf. Flush valve urinals on the market today use 1.0 or 0.5 gpf.

Older urinal valves can be replaced with efficient valves similar to flush valve toilets. To achieve the proper flush pattern, the urinal should also be replaced with a low volume urinal. However, because urinals only flush liquids, poor flush performance is less of a concern. Checking the flush valves bi-annually can reduce the amount of water lost from broken valves that are stuck open, causing a continuous flush. Periodic leak detection at the pipe and handle connections is an important maintenance practice, as well.

No-water urinals can fairly easily replace existing flush valve urinals. However, a thorough economic analysis, including both the initial cost and the long-term O&M costs, is very important. No-water urinals have no valves to maintain, but the sealant in the cartridge must be

Table 5 Estimated consumption and savings by efficient toilets in commercial facilities

Fixture/year of installation	Water use (gpf)	Daily consumption male/female (gal/day)		Annual consumption male/female (gal/year)		Annual water savings—ultra low-flow toilets (ULFs) retrofit male/female (gal/year)		Annual water savings—high efficiency toilets (HETs) retrofit male/female (gal/year)	
HETs	1.0–1.28	1.0–1.28	3.0–3.84	260–333	780–999	—	—	—	
1994–present (ULFs)	1.6	1.6	4.8	416	1,248	—	—	83–156	250–468
1980–1994	3.5	3.5	10.5	910	2,730	494	1,482	577–650	1,732–1,950
Pre-1980	5.0	5.0	15.0	1,300	3,900	884	2,652	967–1,040	2,902–3,120

Fig. 5 Urinal with infrared sensor.

replaced, and some brands require periodic cartridge replacement. Additional training for janitorial and maintenance staff is required.

Other maintenance issues with no-water urinals are still not well understood. Some evidence points to maintenance savings because there may be a decrease in sewer line calcification. With older sewer lines and hard water, uric acid in urine can combine with minerals in water and deposit on the interior of the pipe. Other evidence points to the opposite; no-water urinals can possibly cause line stoppage. At the time of this writing, no major independent study has quantified these issues.

A small demonstration of no-water urinals prior to a large replacement may help ease the adjustment period and solve issues more quickly.

Table 6 Estimated consumption and savings by efficient urinals in commercial facilities

Fixture/year of installation	Water use (gpf)	Daily consumption (gal/day)	Annual consumption (gal/year)	Annual water savings—Energy Policy Act (EPAct) standard retrofit (gal/year)	Annual water savings—high efficiency retrofit (gal/year)	Annual water savings—no-water retrofit (gal/year)
No-water	0	—	—	—	—	—
High efficiency	0.5	1.0	260	—	—	260
1994–present	1.0	2.0	520	—	260	520
1980–1994	1.5	3.0	780	260	520	780
1980–1994	3.0	6.0	1,560	1,040	1,300	1,560
Pre-1980	5.0	10.0	3,900	3,380	3,640	3,900

Auditing

After determining the vintage of the flush valve units and, subsequently, their water consumption per flush, the total water consumption can be estimated based on the number of men in the facility and the average number of uses per day. Daily and annual water use per man for each urinal vintage is summarized in Table 6. The table also shows the savings potential that can be achieved by retrofitting the urinal to a more efficient unit. It is assumed that men use urinals twice per day during an 8-hour work day.

DISHWASHERS

Conventional, standard-size residential dishwashing machines (see Fig. 6) use about 9.5 gallons of water per cycle. More efficient EnergyStar™ machines use less than six gallons per cycle on average.[8] Compact size machines appropriate for smaller loads are also available, and they utilize less water and energy per cycle. In any dishwasher, washing full loads is always most efficient.

Standards currently regulate only energy use in dishwashing machines; however, because about 56% of dishwasher energy use is water heating, conserving energy will most likely conserve water. Federal standards require at least 0.46 cycles per kWh (of machine electrical energy and water heating energy) for standard size machines and 0.62 cycles per kWh for compact machines. EnergyStar™ currently has requirements only for standard-size machines—0.58 cycles per kWh. In 2007, EnergyStar™ will require standard machines to obtain 0.65 cycles per kWh and compact machines to obtain 0.88 cycles per kWh.[9]

Pressure or flow regulators can be installed to limit the water entering the machine in order to match the manufacturer's recommendations.[4] Machines should be checked for proper operation and correct water use.

Auditing

Residential style dishwasher water consumption can be estimated by obtaining the gallons per cycle, found either on the machine or from manufacturer's product data. Use patterns need to be collected via interviews.

CLOTHES WASHERS

A conventional top-loading (vertical-axis) residential machine uses about 40 gallons per load (gpl).[10] Efficient top-loading machines that use a high-pressure spray instead of tub soaking to rinse clothes, and sensors to measure the hot water needs per load are also available. Horizontal-axis machines use gravity to assist in spinning on a horizontal axis, and consume only 15–25 gpl[3,9] because only the bottom third of the machine is filled with water. Horizontal-axis machines are mostly front-loading (see Fig. 7), but top-loading machines are available as well.

Federal standards currently require clothes washers to meet a minimum energy factor (MEF) of 1.04 cubic feet (ft^3) per kWh per cycle, which will be raised to 1.26 ft^3/kWh/cycle in 2007. EnergyStar™ requires 1.42 ft^3/kWh/cycle. The MEF takes water use and dryer energy into account, not just washer energy.[11]

Replacing vertical-axis machines with horizontal-axis machines saves about 40% of the typical water used, plus almost 60% of the typical energy used, not including the dryer energy savings, and about 50% of the typical amount of detergent used.[10] Additional energy savings is achieved as a result of higher-efficiency motors and high spin speeds that result in dryer clothes than with vertical-axis machines.

In apartment buildings or other multi-family residential facilities, common coin-operated laundry instead of individual machines in each apartment has reduced water consumption by about 72%, according to the Multi-Housing Laundry Association study.[4]

Fig. 6 Standard-size residential dishwasher.

In any laundry facility or machine, washing only full loads will reduce overall energy and water use. Using cold water whenever possible saves energy by saving heated water.

Auditing

To determine laundry water use, the frequency of machine use and water consumption per load must be known. Frequency of use will probably have to be estimated using employee or resident interviews. Water consumption may be written on the machine; if not, manufacturer and model information can be used to obtain water use data. Incomplete data can be combined with typical consumption values based on the type of washer, as previously noted.

GENERAL CONSERVATION TECHNIQUES

Education and outreach is a fundamental part of water conservation. A leak of one drop per second can waste 36 gallons of water per day[5]—a maintenance team's quick response can greatly limit this waste. User-friendly mechanisms to report fixture problems in public restrooms can help achieve faster repairs.

Energy is conserved with many water conservation strategies, especially when hot water is saved. It is important to note that lowering the water heater temperature to what is needed, generally no higher than 120°F, is a simple no-cost measure that saves energy without involving water conservation.

Domestic water consumption should utilize water management planning as described in the following entry, "Commercial and Industrial Water-Using Equipment."

CONCLUSION

Water-using technologies are an important aspect of energy engineering because their use has such a wide, but often unrealized, impact on energy. Simple retrofits and proper O&M can have a significant impact on domestic water and energy consumption. Auditing to

Fig. 7 Front-loading horizontal-axis clothes washer.

calculate current consumption provides the baseline from which to start reducing daily water use.

RESOURCES

General

American Water Works Association
www.awwa.org
Information on American Water Works Association (AWWA's) activities, programs, conferences, and publications and useful links

Water Efficiency and Conservation

Environmental Protection Agency Water Efficiency Program
www.epa.gov/owm/water-efficiency/index.htm
Efficiency tips on agriculture, residential water use, drought management, water recycling, and more
California Urban Water Conservation Council
www.cuwcc.org
Statewide water conservation awareness program with urban water conservation Best Management Practices

Water and Energy

State of California Energy Commission Energy and Water Connection
www.energy.ca.gov/process/water/water_index.html
Information specifically tying together energy and water
Watergy
www.watergy.org
A division of the Alliance to Save Energy, drawing the links between water supply and distribution and limited energy resources

Education

Saving Water Partnership
www.savingwater.org

Educational resources about water conservation from Seattle-area utilities

Federal Guidance

Federal Energy Management Program (FEMP)—Water Efficiency

www.eere.energy.gov/femp/technologies/water_efficiency.cfm

Guidance for federal agencies on water efficiency improvements

Energy Center of Expertise—Water Management Guide

www.gsa.gov/gsa/cm_attachments/GSA_DOCUMENT/waterguide_new_R2E-c-t-r_0Z5RDZ-i34K-pR.pdf

A comprehensive handbook for federal facility managers on water conservation

Toilets

Toiletology 101

www.toiletology.com

General information on toilets

How Stuff Works—Tank Toilets

www.home.howstuffworks.com/toilet.htm

Basics of how a tank toilet works

ACKNOWLEDGMENTS

The authors would like to thank Mark Halverson for his technical peer review, Sue Arey for her editorial review, and David Trost and David Gualtieri for their photographic assistance.

REFERENCES

1. http://www.energystar.gov/index.cfm?c=products.pr_protect_water_supplies (accessed December 2005).
2. Rehfeld, B. *Spending. Steps to Take Before the Collective Well Runs Dry*; New York Times: New York, 2005.
3. McMordie Stoughton, K.L.; Solana, A.E.; Elliott, D.B.; Sullivan, G.P.; Parker, G.B. *Update of Market Assessment for Capturing Water Conservation Opportunities in the Federal Sector*, PNNL-15320; Pacific Northwest National Laboratory: Richland, WA, 2005; 38.
4. *A Water Conservation Guide for Commercial, Institutional and Industrial Users*; New Mexico Office of the State Engineer: Albuquerque, NM, 1999; 33–41.
5. *Water Management Guide*; U.S. General Services Administration: Kansas City, MO, 2000; 61–62.
6. http://www.eere.energy.gov/femp/pdfs/watergy_manual.pdf (accessed December 2005).
7. Vickers, A. *Handbook of Water Use and Conservation*; Water Plow Press: Amherst, MA, 2001; 26.
8. Sullivan, G.P.; Elliott, D.B.; Hadley, A.; Hillman, T.C.; Ledbetter, M.R.; Payson, D.R. *SWEEP—Save Water and Energy Education Program*, PNNL-13538; Pacific Northwest National Laboratory: Richland, WA, 2001; 34.
9. http://www.energystar.gov/ia/partners/prod_development/revisions/downloads/dishwashers/ES_Dishwasher_Criteria Letter_finalDec20.pdf (accessed December 2005).
10. http://www.toolbase.org/techinv/techDetails.aspx?technologyID=177 (accessed December 2005).
11. http://www.aceee.org/consumerguide/topwash.htm (accessed December 2005).

Wind Power

K. E. Ohrn
Cypress Digital Ltd., Vancouver, British Columbia, Canada

Abstract
This entry discusses a broad range of topics as an overview of the field of wind power. The topics include current penetration, market shares, technology, costs, governments, and regulation.

INTRODUCTION (WEB PREVIEW)

This article is a broad overview of wind power. It covers a range of topics, each of which could be expanded considerably. It is intended as an introductory reference for engineers, students, policy-makers, and the lay public.

Wind power is a small but growing source of electrical energy. Its economics are well known; there are several large and competent manufacturers, and technical problems are steadily being addressed. Wind power can now be considered as a financially and operationally viable alternative when planning additional electrical capacity. However, as shown below, wind power is only a minor component of present energy sources.

World final energy consumption (2002)[1]: 100%
World final electrical consumption (2002)[2]: 16.1%
World final wind power consumption (2004)[3]: 0.15%

Wind power's main deficiency as a power source is variability. Since wind velocity cannot be controlled or predicted with pinpoint accuracy, alternatives must be available to meet demand fluctuations.

Wind power carries few environmental penalties and makes use of a renewable resource. It has the potential to become a major but not dominant part of the future energy equation.

History

People have used wind to move boats, grind grains, and pump water for thousands of years. Wind-powered flour mills were common in Europe in the 12th century. In the 1700s, the Dutch added technical sophistication to their windmills with improved blades and a method to follow the prevailing wind. Isolated farms in the last century used windmills to generate electricity until the availability of the electrical power grid became widespread.

Keywords: Wind power; Wind energy; Electrical energy; Environment; Wind farm; Wind turbine; Renewable.

Past interest in wind power has tended to rise and fall with fuel prices for the predominant method of electrical production—thermal plants burning oil, natural gas, and coal.

CURRENT

Electrical Production

Although small, wind power is a fast-growing part of the energy picture. Since 1990, worldwide installed capacity has grown about 27% (Table 1).[3]

Business is good for the leading manufacturers of wind power devices. Sales have increased; the technology is stable and predictable, with low maintenance and high availability.

Geographical Distribution (Countries)

The European Union had around 72% of installed capacity in 2004, and Germany, Spain, and the United States accounted for 66.1%. Denmark, Spain, and Germany had by far the largest 2004 capacity in terms of MW per million populations, and 10.6% of the world's population had 81.9% of its wind power capacity. In 2004, Denmark produced about 20% of its electrical power from wind power in 2004 (Table 2).[3]

Manufacturing capacity in 2002 was largely confined to this group of countries, with five big vendors accounting for 76% of sales. European Union vendors accounted for 85% of manufacturing market share.[3]

ECONOMICS

Cost per kWh

Wind power is a viable method of producing electricity that is capital intensive with low operating costs. The cost of production compares favorably with traditional fossil fuel or nuclear plant costs.

Table 1 Capacity growth

Year	Capacity (MW)	Growth rate year-over-year (%)
1990	1,743	13.8
1991	1,983	17.0
1992	2,321	20.7
1993	2,801	26.1
1994	3,531	36.5
1995	4,821	26.6
1996	6,104	25.1
1997	7,636	33.0
1998	10,153	33.9
1999	13,594	27.7
2000	17,357	66.3
2001	28,857	7.9
2002	31,128	26.9
2003	39,500	20.3
2004	47,500	13.8
	Average Growth Rate	26.7

Source: Reprinted with permission from European Wind Energy Association (see Ref. 3).

The major cost elements of a modern wind power installation are as follows[3,5,9]:

Capital
Onshore: 1200–1500 USD/kW
Offshore: 1700–2200 USD/kW
Operating: usually about 1.5%–2.0% of capital cost per year.[12]

Capital costs include wind capacity survey and analysis, land surveying, permits, roads, foundations, towers and turbines, sensors and communications systems, cabling to transformers and substations, maintenance facilities, testing, and commissioning. By far the largest individual capital cost is the turbine (up to 75%).

Operating and maintenance costs include management fees, insurance, property taxes, rent, and both scheduled and unscheduled maintenance.

Financing costs are a major portion of energy production costs, making them very sensitive to interest rates, incentives and subsidization.

Energy production cost estimates vary considerably. Optimists in the industry, such as the British Wind Energy Association, quote a low of 4.8 USD cents per kWh for an onshore plant in an optimal location. Pessimists, like the Royal Academy of Engineering in the UK, quote up to 13.2 USD cents per kWh for an offshore plant, partly by including a controversial 3.1 USD cents per kWh cost for

Table 2 Capacity distribution

Country	Wind power installed capacity (MW) 2001	2002	2003	2004	One-year (%)	Three-year (%)	Population (millions)	Capacity (MW/Million)	Percent of world capacity
Denmark	2,456	2,880	3,076	3,083	0.2	7.9	5.4	570.9	6.4
Spain	3,550	5,043	6,420	8,263	28.7	32.5	40.3	205.0	17.2
Germany	8,734	11,968	14,612	16,649	13.9	24.0	82.4	202.1	34.7
Netherlands	523	727	938	1,081	15.2	27.4	16.4	65.9	2.3
USA	4,245	4,674	6,361	6,750	6.1	16.7	295.8	22.8	14.1
Italy	700	806	922	1,261	36.8	21.7	58.1	21.7	2.6
UK	525	570	759	889	17.1	19.2	60.4	14.7	1.9
Japan	357	486	761	991	30.2	40.5	127.4	7.8	2.1
India	1,456	1,702	2,125	3,000	41.2	27.2	1080.4	2.8	6.3
China	406	473	571	769	34.7	23.7	1306.4	0.6	1.6
Total	22,952	29,329	36,545	42,736	16.9	23.0	3073		89.2

Source: From Global Wind Energy Council—"Wind Force 12" (see Ref. 17).

Table 3 Wind power costs

Wind power costs	Cents (US) per kWh			
	Wind onshore	Wind offshore	Coal CFB	Gas CCGT
RAE [15]	6.8	9.9	4.8	4.0
RAE 2[a]	10.1	13.2	9.2	6.0
EWEA [3]	5.5	8.5		
AWEA (5)	6.6			
AWEA 2[b]	5.4			
BWEA (minimum) [4]	4.8	6.9	4.8	4.8
BWEA (maximum)[c]	6.8	9.1	6.8	5.9
Euro to USD conversion (2004)				1.22
GBP to USD conversion (2004)				1.83

[a]Adds 3.1 cents per kWh for wind power backup capacity and 1.9–4.6 cents/kWh for coal and gas carbon capture.
[b]AWEA figures adjusted to delete 1.8 cents US/kWh production tax credit and are for onshore sites with different average wind speeds.
[c]BWEA figures for a range of site types in November 2004 and 1–2 cents/kWh for carbon capture.
Source: RAE, Royal Academy of Engineering (see Ref. 15); BWEA, British Wind Energy Association (see Ref. 4); EWEA, European Wind Energy Association (see Ref. 3); AWEA, American Wind Energy Association (see Ref. 5).

"standby capacity" required to supply demand when wind power is not available (Table 3).

Capacity Factor

The power generated by a wind turbine depends on the speed of the wind, and on how often it is available. At any given site, this is measured by the capacity factor, or the ratio of actual generated energy to the theoretical maximum. Wind power turbine electrical output rises as the cube of the wind speed. When wind speed doubles, energy output increases eightfold. A typical turbine begins to turn when wind speed is at 9 MPH and will cut out at 56 MPH for safety reasons.

Capacity factors vary by site but are typically in the range of 20%–30% with occasional very good offshore sites reaching 40%. The yearly energy output from a wind farm is given by the following formula:

$$\text{Output/year(kWh)} = [\text{Capacity(kW)}] \times [8760\ \text{hours/year}] \times [\text{Capacity factor}]$$

The Site

Power production costs, site size, site design, and energy output and variability will depend mainly on details about the wind. These details include wind speed, wind direction, and the geographical distribution of favorable wind profiles. During analysis of potential sites, most planners use high (60 m plus) anemometer towers—often several of them—to gather at least one year's data per site. These data are usually correlated with national meteorological observations, if these are available and suitable. If not available, it would be prudent to gather site data for a longer period of up to three years.

Investors and regulators are increasingly aware of the crucial nature of wind data in estimating the quantity and timing of potential power production at a specific site. This research is crucial to the financial analysis of a potential wind power venture.

Other site analysis factors are accessibility via road for heavy equipment, electrical grid proximity and capacity, land ownership, and environmental impact.

LOCATION

Favored Geography

Many countries have developed wind charts of broad areas based on meteorological data gathered for weather and aviation purposes. These charts show potential areas for investigation, where wind strength is high and constant over long periods of the year. Once potential sites, and their extent, have been identified, on-site data measurements provide the basis for analysis and modeling of potential energy production for a specific site.

After wind modeling, the site's geographical, environmental, financing, and ownership issues can be explored in detail.

Generally, sites are either onshore or offshore. Onshore sites are cheaper to construct, but have lower capacity factors due to wind turbulence from nearby hills, trees, and buildings. Offshore sites can have more potential energy available due to higher wind speeds and lower turbulence, which also reduces turbine component wear. Good offshore sites can be near high-demand load areas such as coastal cities, which also increases transmission options. Aesthetic and noise concerns are often fewer offshore, and sea-bed environmental concerns can be lower than land-use concerns for an onshore site.

Sizing a Location

Wind farm towers are usually spread over a large area in order to minimize wake losses. A spacing of five rotor diameters is often recommended. In a typical wind farm, the land physically occupied by tower foundations, buildings, and roads is often less than 2% of overall land area.[6] The remaining land is quite suitable for agriculture and other uses.

Limits to Maximum Production

How much capacity exists to generate electricity from wind? Is it possible that we will require more energy than wind can provide? After surveying wind patterns in the United States and applying energy density and extraction calculations, Elliott and Schwartz[14] concluded in 1993 that 6% of the available U.S. land mass could provide 150% of then-current U.S. electrical consumption. Furthermore, the needed land would be sparsely affected by the wind farm installation, with the vast majority of it (95%–98%) unoccupied by tower foundations, roads, or ancillary equipment and suitable for farming, ranching, and other uses. This study excluded land that is environmentally or otherwise unsuitable, such as cities, forests, parks, wildlife refuges, and environmental exclusion areas.

In the European Union, potential wind power capacity is also larger than current electrical consumption.

FUTURE

Projected Growth

Thanks to increasing concern over the environmental effects of greenhouse gas emissions, the rising cost of fossil fuels, and the impending decrease in availability of oil and natural gas, wind power has a bright future.

Current 25%–30% growth rates are likely not sustainable, due to equipment production volume constraints and limits to perceived need for further capacity. Given the Eurocentric, highly clustered nature of current installed capacity, there is significant potential for high-rate growth elsewhere. However, even in European countries like Germany and Denmark, steady growth will be driven by predicted cost reductions in the 10%–20% range and by regulatory and governmental initiatives aimed at reducing emissions from electrical energy production and transitioning to renewable resources.

Projected Cost

Wind power technology is well down the cost improvement curve, with costs having fallen to present levels, below ten cents USD per kWh, from over $1.00 U.S. in 1978. Costs for a medium-sized turbine have dropped 50% since the mid-1980s, reflecting increasing maturity in the market. Cost projections range from a further 9%–17% drop as installed capacity doubles in the near-term future.

Projected Production

With increasing governmental policy support and commitment, growth rates of 15%–20% appear achievable in five to ten years. But there is likely an upper limit to the amount of electrical energy that can be produced from the wind.

Reaching Maximum Production

Production limits for wind power are based on its variable nature. Other types of electrical production capacity will be needed to provide base-load electrical capacity in the event that there is little wind available. Wind power will then become one player in a mix of generating technologies.

The Hydrogen Economy

As wind power becomes a larger portion of electrical supply, occasionally its supply will exceed demand. Rather than simply curtailing wind plant production, it is attractive to think of using this excess electrical power capacity to generate hydrogen via electrolysis. This has the effect of storing wind energy that would otherwise not be harvested. This energy, in the form of hydrogen, can be used directly as a non-polluting fuel or as an input source to fuel cells to produce electricity at a later time.

When there is a significant hydrogen economy, with transmission lines, storage, and fuel cell capacity, this use for wind power will become a very attractive scenario.

Other Issues for the Future

- Learning more about wind and forecasting—predicting the best locations, wind farm output, gusts, and directional shear.[10] This will help reduce financing costs when wind power plant output and impact on the grid are better understood and more predictable.

- Improving the control of demand through incentives around end-user load shedding, rescheduling and simple conservation methods. This could be used to offset wind power production shortfalls as an alternative to other forms of generation.
- Advancing aerodynamics specific to wind turbine blades and control systems.
- Designing extremely large wind tunnels to study wake effects minimization, structural load prediction, and energy output maximization at lower wind speeds.
- Enhancing power system capacity planning models to include wind farm components.
- Re-planting, or upgrading older mechanical and electrical components at existing wind farms.
- Wind farm siting further offshore and on floating platforms.
- Combining wind power and hydroelectric capacity by using surplus wind power to pump water behind dams and so store power that might otherwise be wasted.
- Determining how and whether to allocate full costs of environmental impact to fossil and nuclear plants.

STRENGTHS AND WEAKNESSES

Strengths

Environment

Wind power installations do not emit air pollution in the form of carbon dioxide, sulfur dioxide, nitrogen oxides, or other particulate matter such as heavy metal air toxins. Wind power installations do not use water or discharge any hazardous waste or heat into water. Conventional coal, oil, and gas electric power plants produce significant emissions of all kinds. Nuclear power plants produce dangerous and long-lasting radioactive waste. Greater use of wind power means less impact on health and the environment, particularly regarding climate change due to greenhouse gas emissions.

Renewable

Wind power produces energy from a resource that is constantly renewed. The energy in wind is derived from the sun, which heats different parts of the earth at different rates during the day and over the seasons. Unlike fossil and nuclear plants, the source of energy is essentially inexhaustible.

Costs

Wind power's costs are well known and are dropping to the point at which this technology is very competitive with other means of production. Fuel costs are nil, meaning that fuel costs have no uncertainty. Wind power costs should be more stable and predictable over the lifetime of the plant than power costs for fossil fuel plants.

Local and Diverse

Wind power plants provide energy source diversity and reduce the need to find, develop, and secure sources of fossil or nuclear fuel. This reduces foreign dependencies in energy supply, and reduces the chances of a political problem or natural disaster interfering with and diminishing the supply of electricity.

Quick to Build, Easy to Expand

Wind power plants of significant capacity can be constructed and installed within a year, a much shorter time than conventional plants. The planning time horizon is similar to conventional plants, given the need to accurately survey site wind characteristics and deal with normal environmental and related site issues. This means that capacity can be increased in closer step with demand than with conventional plants. With the right site and design, a wind power plant can be incrementally expanded very quickly.

Weaknesses

Natural Variability

A single wind farm produces variable amounts of energy, and its output is not yet as predictable as a traditional plant. As the geographical distribution and number of wind plants increases, and as research into predicting wind continues, these problems should be minimized, allowing cost-effective and orderly scheduling and dispatch of total grid capacity sources—but it is difficult to see traditional power sources disappearing entirely.

Connection to Grid

As the amount of electrical power supplied by wind power plants increases, concern increases over its effects on the electrical grid.

In order to maintain a reliable supply of electricity that matches demand, utility operators maintain emergency reserve capacity in order to deal with plant outages (failures) and unexpected demand across their entire system. This reserve is in the form of purchased power, unused capacity at conventional plants running below their maximum, or quick-start plants such as gas-fired turbines. Often, conventional plants on the grid are allocated a cost to cover this reserve based on their capacity (large plant, large reserve) and reliability (more outages, more reserve).

The industry is working on ways to determine and allocate this reserve cost for wind power plants. Yet to be agreed upon is the statistical basis for calculating such wind power plant reserve costs. Improvements in day-ahead wind forecasting will greatly reduce the uncertainty around wind plant output, and so decrease the cost burden to provide this reserve.

Several current estimates prepared for U.S. utilities show this reserve cost burden (or ancillary services cost) to increase with the amount of capacity provided by wind power, and to be in the range of 0.1–0.5 cents USD per kWh for penetrations between 3.5 and 29%. In no case was it thought necessary to allocate a reserve equal to 100% of the wind power capacity.[11] German experience is similar,[8] with no additional reserve capacity required for the 14% wind energy share of the national electrical consumption forecast for 2015.

When wind power supplies less than 20% of electrical consumption, these problems are not severe. At larger penetrations, reserves become a major issue. Interestingly, wind power plants may be subject to shutdown or voluntary power reductions in the event of coincident high wind, low demand situations. This is occasionally the case today in Denmark and Spain.

In some cases, wind power sites are situated far from the location of high electrical power demand, placing strain or potential overload on existing transmission facilities. In these cases, there are often cost, ownership and responsibility issues yet to be resolved.

Power quality problems around power factor, harmonic distortions, and frequency and voltage fluctuations are being successfully addressed in modern large production wind farms.[8]

This is one of the most difficult sets of issues facing the future of wind power as it matures from small-scale and local to large-scale penetration.

Local Resource Shortage

In a few places, high-quality wind power sites are not available or are already in production, leaving these places to import electrical power or use traditional sources.

Noise

Noise levels have decreased and are now confined to blade noise in modern units. Generator and related mechanical noise has been effectively eliminated. Noise, however, will always be a significant factor. Blade noise is described as a "whoosh, whoosh" sound, and is in the 45–50 decibel range at a distance of 200–300 m. This noise level is consistent with many national noise level regulations. However, this noise buffer zone adds to the overall land requirement for a wind power plant and so increases costs.

Visual Impact

Onshore wind farms are highly visible due to the height of towers and the size of the blades and generator. The impact of this varies with each person. Each wind plant operator needs to determine the levels of support and opposition from those who live and work within sight of the plant. Offshore plants attract fewer detractors than onshore plants—one of the reasons for their increasing popularity.

Offshore wind plants are less likely to cause unwanted noise since they are far from human habitation. This reduces turbine and blade design constraints and can lead to higher capacity factors.

Bird Impact

Bird deaths are a regrettable reality. The bird death rate at a specific wind farm project is quite variable. Several early wind farms (Altamont Pass, California, and La Tarifa, Spain) caused concern over death rates. The California Energy Commission estimates the death rate at Altamont (5400 turbines) to be 0.33–0.87 bird deaths per turbine per year.[16] The overall recent U.S. national average[13] is 2.3 bird deaths per turbine per year. Prudently located sites are off migration routes and not in nesting, over-wintering, or feeding areas. Their tower designs do not offer nesting or even roosting places. In such locations, death rates are lower, and overall impact is much lower than that caused by other types of human activity.

Since climate change is a very serious environmental problem faced by bird populations, wind power and other renewables are an important part of the solution.

TECHNOLOGY

Overview

Wind turbine design has three major components, and there are large economies of scale in design.

- Tower height: Wind turbine energy output is proportional to the cube of wind speed. Since moving air (wind) is subject to drag and turbulence from its contact with the earth and the objects on the earth, wind speed increases with height (vertical shear). The higher the tower, the more advantage there is for power generation. The tradeoff is between tower costs and increase in power generation. Typically, tower heights are rising, and are currently in the 100 m range. Off shore, vertical wind shear is generally less than onshore, so towers can be shorter, with wave height clearance being the factor that determines tower height (Fig. 1).
- Blade diameter: The power capacity (watts) of a wind turbine varies with the square of its blade diameter, because a blade with a larger diameter has a larger area available for harvesting the wind energy passing through it. The coefficient of performance defines the actual power capacity compared to the

Fig. 1 Typical large wind turbine. Note entrance steps and utility pole at base for scale. (Photo courtesy of Suncor Energy, Inc.)

maximum—how much energy can be extracted from the wind compared to the available energy. Modern wind turbines can achieve a coefficient of performance approaching 0.5, very close to the theoretical maximum of 0.59 derived by Betz.[3,18] This maximum is derived from the concept that if 100% of wind energy were extracted, the wind exiting the turbine would be at zero speed, so no new air could enter the turbine. Larger capacity turbines benefit significantly from economies of scale in foundation and support costs as well as swept area (Fig. 2).

- Controls and generating equipment: The turbine's hub (or nacelle) is the most costly component and contains the generator, gear boxes, yaw controls, brakes, cooling mechanism, computer controls, anemometer, and wind directional vane.

Generators

As the blades turn, they drive a generator to produce electricity. Generating capacity ranges from a few hundred

Fig. 2 Site assembly of large wind turbine nacelles and blades. (Photo courtesy of Suncor Energy, Inc.)

kW to over 3 MW. In older designs, there is usually a 40:1 gearbox to match low, fixed rotor speeds (~30 RPM) to required generator speed (1200 RPM for a 60 Hz output, 6-pole generator). The gearbox often incorporates brakes as a part of the overall wind turbine control system. Generators that operate at low RPM are available and are called direct drive generators. These would eliminate the gearbox.

In more modern designs, rotor speed is variable and controlled to optimize power extraction from the available wind. Generator output is converted to d.c. and then back to a.c. at the required grid frequency and voltage. The conversion equipment is sometimes located at a central part of the wind power plant.

This is an active area for ongoing technical innovation.

Blades

In order to maximize power capacity through size, blades are very long, up to 50 m. To minimize noise, they must turn slowly so as to reduce tip speed, the primary blade noise source. Typical rotation speeds are in the 10–30 RPM range. Blades are increasingly made from composites (carbon fibre reinforced epoxy resins).

Rotor blade aerodynamics[19] have much in common with the aerodynamics of a propeller or a helicopter blade, but they are sufficiently different that the aerodynamics of wind turbine blades is an evolving field. The difference is that wind turbine airflows are unsteady due to gusting, turbulence, vertical shear, turbine tower upstream shadow, yaw correction lag, and the effects of rotation on flow development. For example, at present it can be difficult to predict rotor torque (and therefore power output) accurately for normal turbine operating conditions. Further development of theory and modeling tools should allow the industry to improve rotor strength, weight, power predictability, power output, and plant longevity while controlling cost and structural life.

For a given site wind velocity, the rotor blade's tip has very different air flow than its root, requiring the blade to be designed in a careful twist. The outer third of the blade generally produces two-thirds of the rotor's power. The third nearest the hub provides mechanical strength to support the tip, and also provides starting torque in startup situations.

Each blade generally has lightning protection in the form of a metallic piece on the tip and a conductor running to the hub.

Some manufacturers place Whitcomb winglets at the blade tips to reduce induced drag and rotor noise, in common with aircraft wing design.

In order to control blades during high wind speeds, some are designed with a fixed pitch that will progressively stall in high wind speeds. Others incorporate active pitch control mechanisms at the hub. Such control systems use hydraulic actuators or electric stepper motors and must act very quickly to be effective.

Wind Sensors

Wind turbines incorporate an anemometer to measure wind speed and one or more vanes to measure direction. These are primary inputs to the control mechanism and data gathering systems usually incorporated into a wind turbine.

Control Mechanisms (Computer Systems)

Control systems are used to yaw the wind turbine to face into the wind, and in some designs to control blade pitch angle or activate brakes when wind gets too strong. In sophisticated cases, the controllers are redundant closed-loop systems that operate pumps, valves and motors to achieve optimum wind turbine performance. They also monitor and collect data about wind strength and direction; electrical voltage, frequency and current; nacelle and bearing temperatures; hydraulic pressure levels; and rotor speeds, vibration, yaw, fluid levels, and blade pitch angle. Some designs provide warnings and alarms to central site operators via landline or radio. Manufacturers do not release much detail about these systems, since they are a critical contributor to a wind turbine's overall effectiveness, safety, and mechanical longevity.

ROLE OF GOVERNMENTS AND REGULATORS

Governments play a large part in determining the role and scale of wind energy in our future mix of energy production capabilities.

Subsidies, Tax Incentives

As part of programs to encourage wind power production, the following are used in varying ways[7]:

- Outright subsidies, grants and no-interest loans.
- Tax incentives such as accelerated depreciation.
- Fixed prices paid for produced electrical power.
- Renewable energy quantity targets imposed on power utility operators.

Grid Interconnection and Regulatory Issues

Since many power utility operators are owned by governments, and most are regulated heavily, governments have a role to play in encouraging solutions to grid interconnection issues. There must be a political will to address issues, find solutions, and develop practices and different management strategies that will allow greater penetration of wind power into the electrical supply.

Improving Wind Information

Climate and environmental information is most often collected and supplied by national governments in support of weather and aviation services. Wind atlases are an invaluable resource to the wind plant planning process. National efforts to improve long- and short-term wind forecasting, atmospheric modeling tools, and techniques will benefit wind power projects' ability to forecast power output for long-term and short-term planning purposes.[10]

Environmental Regulation

In this controversial area, government can tighten its regulation of air quality, carbon emissions and other environmental areas. This would have the effect of increasing the apparent cost of conventional thermal electrical power, which is responsible for significant emissions. It is often argued that wind power would already be cost competitive if environmental and health costs were to be fully allocated to conventional oil, gas, and particularly coal-powered plants, or if such plants were required to make investments to significantly reduce emissions.

CONCLUSION

Western societies depend on a steady supply of energy, much of it in the form of electricity. Most of that supply comes from thermal plants that burn oil, natural gas, and coal, or from nuclear plants. Where will our electrical energy come from in the future? How will we keep our environment livable and healthy?

One part of this answer lies in wind power. Its costs are within reason; the technology has matured with some gains yet to be realized; it carries little environmental penalty; and the source of its energy is renewable. As long as the sun heats the earth, wind power will be available.

Wind power will not likely be the complete answer; it is an intermittent source because the wind doesn't always blow. But there is a very large amount of it available for us to harvest. As wind power moves quickly from small-scale to large-scale, its future path depends on governments and regulators as much as it does on technical innovators and manufacturers.

REFERENCES

1. Annual Energy Review 2003, DOE/EIA-0384(2003), Energy Information Administration, U.S. Department of Energy, Washington, DC, September 2004.
2. Key World Energy Statistics - 2004 Edition, International Energy Agency, OECD/IEA - 2, rue André-Pascal, 75775 Paris Cedex 16, France or 9, rue de la Fédération, 75739 Paris Cedex 15, France, 2004.
3. Wind Energy, The Facts, Analysis of Wind Energy In the EU-25, European Wind Energy Association, Brussels, Belgium, 2004.
4. BWEA Briefing Sheet, Wind and the U.K.'s 10% Target, The British Wind Energy Association, http://www.bwea.com/energy/10percent.pdf (accessed July 2005).
5. The Economics of Wind Energy, American Wind Energy Association, http://www.awea.org/pubs/factsheets/EconomicsOfWind-Feb2005.pdf (accessed July 2005).
6. Saddler, H., Dr.; Diesendorf, M., Dr.; Denniss, R.; A Clean Energy Future For Australia, A Study By Energy Strategies for the Clean Energy Future Group, Australia, March 2004.
7. Policies to Promote Non-hydro Renewable Energy in the United States and Selected Countries, Energy Information Administration, Office of Coal, Nuclear, Electric and Alternate Fuels, United States Department of Energy, Washington, DC, February 2005.
8. BRIEFING May 10th, 2005. German Energy Agency DENA study demonstrates that large scale integration of wind energy in the electricity system is technically and economically feasible. http://www.ewea.org/documents/0510_EWEA_BWE_VDMA_dena_briefing.pdf (accessed July 2005).
9. Offshore Wind Experiences, International Energy Agency, 9, rue de la Fédération, Paris Cedex 15, 2004.
10. Milborrow, D., Forecasting for Scheduled Delivery, Windpower Monthly, December 2003.
11. Utility Wind Interest Group—Wind Power Impacts On Electric-Power-System Operating Costs, November 2003.
12. Danish Wind Industry Association, http://www.windpower.org/en/tour.htm (accessed July 2005).
13. Wind Turbine Interactions With Birds and Bats: A Summary of Research Results and Remaining Questions, Fact Sheet, Second Edition, National Wind Coordinating Committee, Washington, DC, November 2004. http://www.nationalwind.org/publications/avian/wildlife_factsheet.pdf (accessed July 2005).
14. Elliott, D.L.; Schwartz, M.N., Wind Energy Potential In the United States, Pacific Northwest Laboratory, PNL-SA-23109, Richland, Washington, USA, September 1993.
15. The Cost of Generating Electricity, Royal Academy of Engineering, http://www.raeng.org.uk/news/publications/list/reports/Cost_of_Generating_Electricity.pdf (accessed July 2005).
16. Developing Methods to Reduce Bird Mortality in the Altamont Pass Wind Resource Area, August 2004, http://www.energy.ca.gov/reports/500-04-052/500-04-052_00_EXEC_SUM.PDF (accessed July 2005).
17. Wind Force 12, Global Wind Energy Council, June 2005, http://www.ewea.org/03publications/WindForce12.htm (accessed July 2005).
18. Proof of Betz Theorem, Danish Wind Industry Association, http://www.windpower.org/en/stat/betzpro.htm (accessed July 2005).
19. National Wind Technology Center, http://www.nrel.gov/wind/about_aerodynamics.html (accessed July 2005).

Window Energy

William Ross McCluney
Florida Solar Energy Center, University of Central Florida, Cocoa, Florida, U.S.A.

Abstract
Modern factory-built windows and window components are available which can greatly reduce that portion of building energy consumption attributable to the windows. The adverse energy consequences of using larger window apertures for good daylighting (thereby realizing its attributes) can be reduced or even eliminated. The result can be buildings with greater comfort, more productive occupants, and substantially reduced purchased energy costs. As we pass the peak of world oil production and approach peaks in other fossil-fuel supplies, considerable energy and dollar savings are possible through the use of the new high performance windows.

INTRODUCTION

Windows have been included in building envelopes since the first buildings were constructed—mainly for the illumination and view that they provide. The addition of glazing provided increased protection from wind and rain. In the past, windows were considered "holes in the insulation." As a result, window areas have generally been kept modest in size, especially in buildings designed for hot or cold climates. Modern window designs and materials have made it possible to minimize—and, in some cases, avoid altogether—the energy costs of windows, enabling builders to utilize larger areas for aesthetic, viewing, and health benefits. In some circumstances, good windows can outperform opaque insulated walls, energy-wise.

The energy performance of a building's window system can be estimated through a variety of engineering calculations. Due to the vagaries of weather and climate, these measures are best performed with the aid of computer simulation, wherein the hourly energy and illumination performances of windows can be assessed quickly with modest precision. In window design, care must be taken to ensure strong energy performance sans excessive glare, overheating, or thermal discomfort.

This article examines the fundamentals of window energy performance and presents some issues surrounding window and shade design and product selection. A separate chapter discusses daylight design more generally.

FUNDAMENTAL QUANTITIES

Radiant flux, Φ, is the time rate of radiant energy flow (unit: watt).

Irradiance, E, is the radiant flux per unit area at a specified point in a specified surface that is incident on, passing through or emerging from that point in the surface, or the average of this quantity over the surface's area (units: watt m^{-2}).

Conductive heat flow, Q, is the flow of heat through a specified distance in a material per unit area and per unit temperature difference. It has units of energy flux per unit area and per unit temperature difference (units: $W\,m^{-2}\,K^{-1}$).

An important characteristic of radiant flux is its distribution over the electromagnetic spectrum, known as a spectral distribution or spectrum. The Greek symbol λ is used to symbolize the wavelength of monochromatic radiation, defined as radiation with only one frequency and wavelength. The symbol for frequency is the Greek symbol ν. The relationship between frequency ν and wavelength λ is shown in the equation

$$\lambda\nu = c \tag{1}$$

where c is the speed of propagation in the medium (more familiarly, the "speed of light," even though "light" is properly applied only to the visible portion of the spectrum). The spectral concentration of radiant flux at a given wavelength λ is given the name "spectral radiant flux" and represented by the symbol Φ_λ (units: $W\,nm^{-1}$). The wavelength dependence is represented by the functional notation, thus, $\Phi_\lambda(\lambda)$, where the subscript denotes the concentration of the quantity at a specific wavelength.

Spectral irradiance is the spectral "concentration" of the irradiance E, and its wavelength dependence is denoted by the symbol $E_\lambda(\lambda)$ (units: $W\,m^{-2}\,nm^{-1}$).

Names for the relevant regions of the electromagnetic spectrum are provided in Table 1.[1]

ENERGY BASICS

Heat Transfer

Energy is transferred through windows by the three mechanisms: radiation, conduction, and convection, illustrated in Fig. 1.

Win–Wire

Keywords: Windows; Daylight; Glazing; Aperture; Illumination; View; Energy.

Table 1 CIE vocabulary for spectral regions

Name	Wavelength range
UV-C	100–280 nm
UV-B	280–315 nm
UV-A	315–400
VIS	Approx. 360–400 to 760–800 nm
IR-A[a]	780–1,400 nm
IR-B	1.4–3.0 μm
IR-C[b]	3 μm–1 mm

[a]Also called "near IR" or NIR.
[b]Also called "far IR" or FIR.
Source: Excerpts from International Lighting Vocabulary (see Ref. 1).

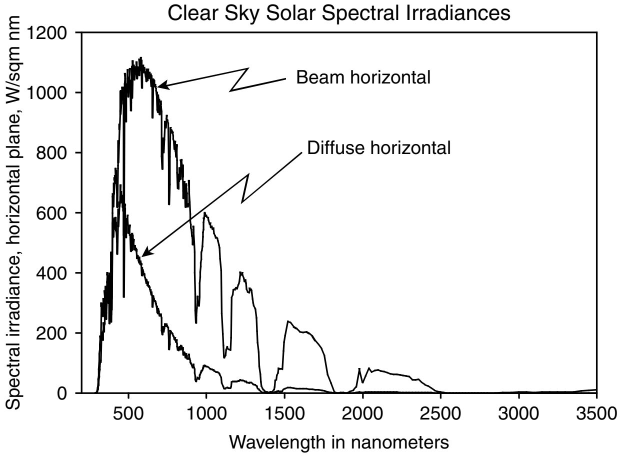


Fig. 2 Spectral irradiance distributions for clear sky direct beam and diffuse sky radiation on a horizontal plane.

Radiation Heat Transfer

The primary form of radiative heat transfer through windows is solar energy. Irradiance spectra for the sun and the diffuse sky for a clear day are shown in Fig. 2.

All objects above an absolute zero temperature emit radiation. Those with temperatures much cooler than the sun emit radiation in the far IR (FIR) region. Radiative transfer can also occur in windows over this spectral band. The FIR region is separated significantly from that of the solar spectrum.[2]

The spectral separation between solar and FIR radiation makes the greenhouse effect possible, a process whereby solar radiation is transmitted effectively through clear glass, but the longer-wavelength radiation emitted from objects receiving the transmitted solar radiation does not transmit back through the glass easily. This mechanism enables greenhouses to remain warm on cold winter days when they receive ample solar radiation, without the need for additional heating. Window manufacturers have learned to take advantage of this spectral separation to improve the winter day and night performance of glazing systems for a variety of applications. When it is cold outside and warm inside, significant quantities of FIR radiation can pass from the warm inner glazing across the gas gap in a double pane window to a colder outer pane. Several mechanisms, described below, are available for reducing the magnitude of this component of heat loss through windows.

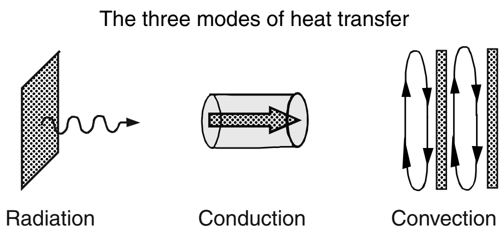


Fig. 1 The three modes of heat transfer.

Conduction Heat Transfer

Heat is transferred via conduction through material that has at least two surfaces with different temperatures. The heat conducts through the material from the warm surface to the cooler one. The rate of transfer depends upon the magnitude of the temperature difference and the thermal conductivity of the material, k, having units of energy per unit time, per unit area, per unit distance through the material, and per unit temperature difference. For homogeneous materials of specific thickness t, the conductance C of that thickness is a product of the conductivity of the material and the distance t through it. Conductance has units of energy per unit time, per unit area, and per unit temperature difference. Window frames with sections that have lower thermal conductivity are more insulating than those without these thermal breaks. Generally, the thermal resistance R is the reciprocal of the thermal conductance C.

Convection Heat Transfer

A warm surface can transfer some of its heat to the adjacent air by conduction. Once the air in a thin film adjacent to the surface receives this heat, it expands slightly and there is a tendency for the air to rise. This air movement can carry heat to or from the glazing of a single-pane window. If the surface is a vertical one, and if the rising warmed air reaches a boundary (like the window frame around an insulating gas space between the panes of a double pane window), it tends to move laterally. If there is a cooler vertical second pane, the heated air will tend to move downward. As it descends, it will lose some of its heat via conduction into the cooler pane of glass. The higher density of the cooler air will drive it further downward and it will, at the bottom, move laterally from the cooler toward the warmer surface, filling the space left

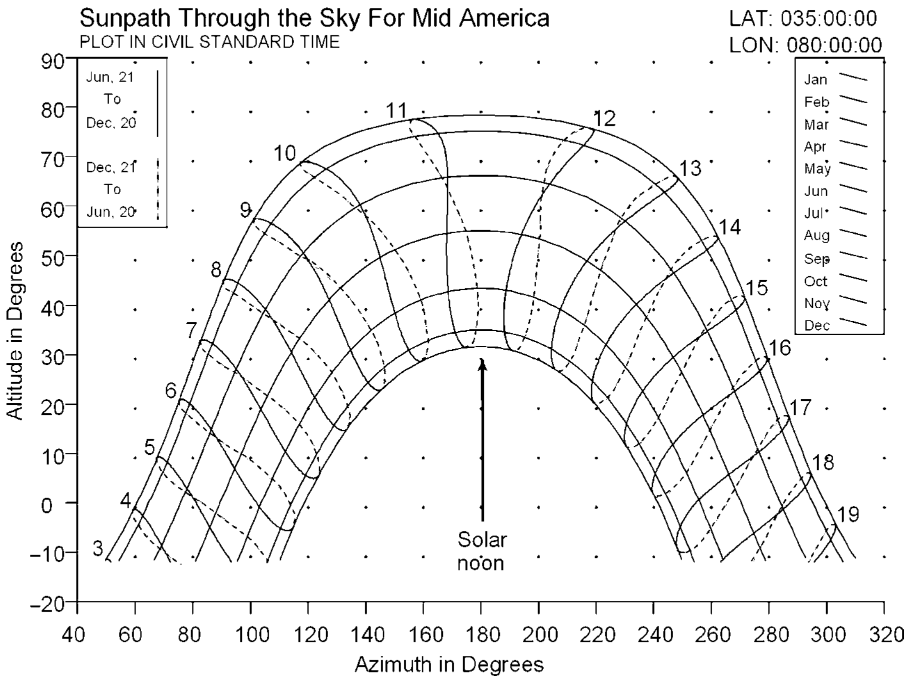


Fig. 3 Sunpath chart for a central U.S. location, in local standard time. Due south is at azimuth 180°. East is at 90°.

by the less dense, rising air next to the warmer pane of glass. This circulation of air from a warmer pane to a cooler one provides the third mechanism for heat transfer through a window—convection.

Air adjacent to a warm pane thus convects to the cooler pane, releases its heat, and repeats the cycle. See Fig. 1. If the air were not moving, it would be a relatively effective insulator since the conductivity of air at atmospheric pressure is quite small. Convective movement of air, however, can "short-circuit" the insulating value of still air and promote heat transfer. Changing the gap width affects the rate of heat transfer by convection. Generally, there is a non-zero optimum gap width for a given gas that maximizes the thermal resistance due to convection.

Double pane windows generally have substantially lower conduction and convection heat transfer. In climates with modest year-round inside/outside temperature differences, single pane windows may be adequate. Multiple pane windows exhibit significantly lower energy costs on buildings in climate regions that experience significant inside/outside temperature differences.

THE PATH OF THE SUN THROUGH THE SKY

The sun moves in a predictable manner each day. During the winter in the northern hemisphere at middle latitudes, it rises south of due east (azimuth 90) and sets south of due west (azimuth 270). In the summer, it rises north of due east and sets north of due west. Plots of solar position vs time on a chart of solar coordinates form what is called a sunpath chart. Fig. 3 shows an example of a sunpath chart for latitude 35 degrees north and longitude 80 degrees west. A variety of sunpath calculators are available on the Web.

ORIENTATION AND SHADING

If the paths of the sun for a given location are known, the building and its windows can be oriented to either capture or prevent the entry of solar heat gain from the direct sun and/or diffuse sky light. Shading devices of many kinds can also be applied to the building to block direct solar beam entry. A variety of devices for exterior application are illustrated in Fig. 4.

Interior shades can also be effective if their solar reflectances on the window-facing side are high, thereby causing them to reflect most of the solar radiation incident back toward—and hopefully through—the window to the outside.

Solar heat loading on cooling equipment can be further reduced when desired by placing buffer spaces like garages, laundry rooms, warehouse space, porches, and utility rooms on the east- or west-facing sides of a building.

When passive solar space heating is desired, moderately large window areas can be added to facades that face directions from which strong direct solar radiation can be expected during certain seasons of the year and times of day. This solar gain can be moderated by judicious building and window orientation and through the use of exterior shading devices and vertical and horizontal fins

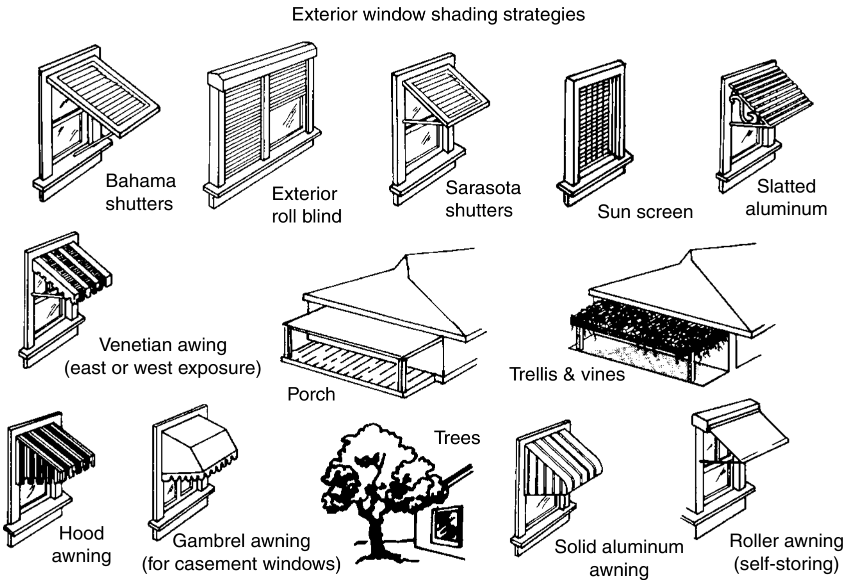


Fig. 4 Window shading options illustrated. Credit: Florida Solar Energy Center, 1679 Clear lake Rd., Cocoa, FL 32922.

and overhangs designed to selectively shade the window when direct sunlight is not desired.

Direct beam entry into a building should be carefully facilitated due to the potential adverse consequences that can result. Bright spots in a generally darker room can be powerful sources of glare and should therefore be avoided or carefully moderated by architectural or interior design features. Localized overheating and the fading of interior furnishings are additional difficulties that can be produced by direct beam entry. Both exterior and interior shades, properly operated through either manual or automatic motorized control, can be effective means of reducing glare and controlling direct beam solar heat gain, while still allowing unobstructed views of the exterior when the sun is not a problem.

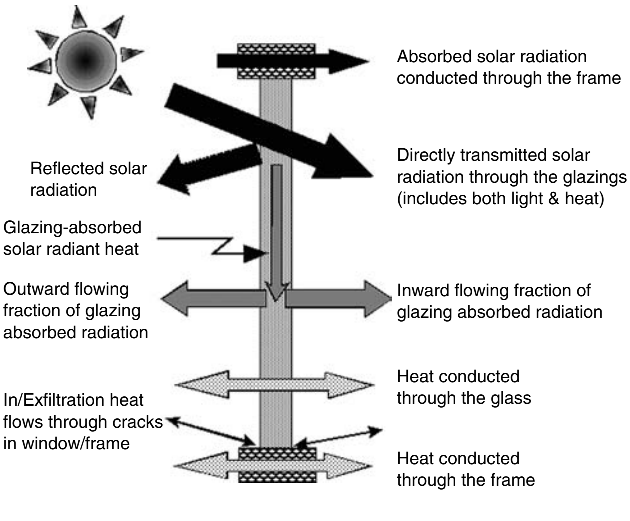


Fig. 5 Heat flows through window glass and window frames.

WINDOW HEAT TRANSFERS

Solar radiant heat is conducted through a window via several mechanisms, depicted schematically in Fig. 5.

The thermal conductivity of glass is typically about $1.5\,\mathrm{Wm^{-1}\,K^{-1}}$. Transparent acrylic and polycarbonate plastics are sometimes used as glazing materials, and their conductivities (around $0.2\,\mathrm{Wm^{-1}\,K^{-1}}$) are somewhat lower than that of glass, but the inherent thermal conductivity of air in the absence of convection is approximately $0.002\,\mathrm{Wm^{-1}\,K^{-1}}$, a much lower value. Double pane windows with a correctly sized gas gap can therefore significantly reduce the conductive heat transfers through them. Use of a more insulating gas, such as Argon or Krypton, can further reduce conductive heat transfer through windows. Due to convection, however, the thermal resistance of the gas gap is somewhat less than the conductance of still air (or a more insulating gas) with the same gap thickness.

All objects above absolute zero emit radiation. Long-wavelength infrared radiation (often termed "thermal" radiation) is therefore another mechanism of heat transfer in windows (the net heat transfer being from a warm to a cold surface). At the temperatures encountered in glazing systems, this radiation is invisible; it is confined to the long-wavelength (FIR) region of the electromagnetic spectrum.

The combined effects of conduction, convection, and radiation across the gas gap between glazing panes, which results from the temperature difference between inside and outside air, are often lumped together in colloquial descriptions and given the name "conductivity." A double pane glazing system has two panes separated around the edge by a spacer bar, a structural material that keeps the panes the correct distance from one another. The spacer bar is often made of metal to ensure its strength. The higher thermal conductivity of metal results in higher conduction of heat through the glass/spacer/glass edge than through the center of the glass/gas/glass combination. Insulated glazing units (IGUs) usually have an edge seal around the outside perimeter of the unit. Sealant is often used between the glass and spacer bar, as well.

To prevent condensation inside sealed IGUs, a desiccant material is usually incorporated into the spacer bar, which contains small openings to allow water vapor in the gas gap to be absorbed by the desiccant inside. Some spacer bars are made of materials with higher insulating value to reduce conduction transfer, or they may contain a thermal break, a section of non-metallic, lower conduction material.

The glazing system is set within a frame. The total window unit, including the glazing and frame, is called a sash. The sash fits into another frame that is connected to the wall and holds the sash in place. Due to its use of different materials, the frame around a glazing conducts heat differently from the glazing. The National Fenestration Rating Council, responsible for rating the energy performance of windows and other fenestrations in the U.S., requires different values for the combined conductive/convective/FIR radiative transfer of heat through the center of a glazing system, through its edge, and through the framing members surrounding the glass.[3] The total thermal conductance of the entire sash is an area-weighted average of the frame, edge, and center of the glass values. Frames made of metal often contain thermal breaks to reduce conduction heat transfer through them.

If a window has cracks or other openings that permit the flow of air through these openings between the inside and outside of the building, the air so transferred can carry heat with it, thereby reducing the thermal integrity of the window. Such convective heat flow is termed infiltration or exfiltration, depending upon whether the air is flowing from outside to inside or vice versa. Modern factory-built windows generally exhibit relatively little infiltration and the heat gains or losses via this mechanism are generally much smaller than other heat transfers through the window. Windows in very cold areas, such as close to the poles of the Earth during the winter, can be exceptions to this rule if they are not very carefully constructed to minimize infiltration transfers.

In hot climates and buildings with air cooling equipment, the infiltration of hot and humid outside air poses a special problem. The air conditioner has to both cool the air and remove some of its moisture. At times, it takes more energy to remove the moisture (called a latent cooling load) than to lower the air's temperature (called a sensible cooling load). Windows, such as "jalousie" windows with large infiltration openings, often contribute large latent and sensible cooling loads to the building. Repairing or replacing such windows would be a useful energy conservation strategy.

THE "U-FACTOR"—CONDUCTIVE HEAT TRANSFER

The American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) uses the term "U-factor" to designate the overall conductive transfer property of a window: the heat transferred through the window by virtue of the temperature difference between the inside and outside air adjacent to the window. The full value for the window as a whole has units of heat flux per unit area and per unit temperature difference. In the *Systeme International* (SI, the international metric system), the units are $W\,m^{-2}\,°K^{-1}$ and in the inch-pound (IP) system, the common unit is $Btu\,hr^{-1}\,ft^{-2}\,°F^{-1}$. As previously mentioned, National Fenestration Rating Council (NFRC) standards divide the U-factor into frame, edge, and center-of-glass components, using an area-weighted average of these for the total product U-factor. The conductive heat loss Q_c for the whole window is the product of the window's U-factor, its area A, and the temperature difference going from inside t_i to outside t_o.

$$Q_c = UA(t_i - t_o) \quad (2)$$

LOW-E COATINGS

In a multiple pane glazing system, it is common to number the glazing surfaces from outside in. Thus, the outer surface in contact with the outside air is surface 1 and the next one is surface 2. Surface 2 faces the first gas gap in the glazing system. It is common in multiple pane systems to place a coating on surface 2. Sometimes, surface 3 is used for the coating. Many such coatings are relatively soft microscopic multilayer structures that would be damaged while cleaning the window. These coatings are placed only

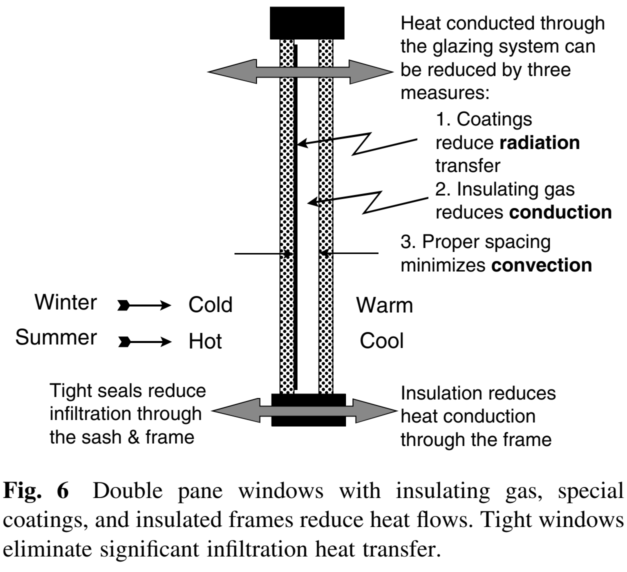


Fig. 6 Double pane windows with insulating gas, special coatings, and insulated frames reduce heat flows. Tight windows eliminate significant infiltration heat transfer.

on one of the inner surfaces in a multiple-pane system, so they can be sealed in the IGU and protected from damage and degradation. Hard coatings with similar properties that can be applied to single pane windows have been developed, but their energy performances are generally not quite as good as for the soft coats. See Fig. 6.

The glass coating has one or two purposes, depending upon its design and construction. The most common purpose is to admit the maximum quantity of solar radiation (over the spectral range from approximately 350–3500 nm) through the glass, while, at the same time, blocking FIR radiation emitted by the warm inner pane from crossing the gas gap and being absorbed by the colder outer pane during cold periods. Such coatings are called low-e coatings because their emissive property over the long-wavelength FIR portion of the spectrum is quite low. Emissivity is the ratio of thermal radiant emission from a surface to the maximum possible emission from the best possible radiator at the same temperature, called a blackbody. Bare glass has an emissivity of approximately 0.94, but low-e coating emissivities from commercial glass companies can be substantially below 0.2. Values as low as 0.09 are common. Such coatings can reduce the emitted long-wavelength radiation from glass by a factor of ten. The effect of this is to reduce the center-of-glass U-factor by reducing the radiative heat transfer component between the panes in the IGU. The U-factor is further lowered by using an insulating gas of the proper thickness between panes.

To distinguish these low-e coatings from a second kind of coating (to be described), they are given the generic name high solar gain low-e coating. An alternative name that has been used is “cold climate low-e coating.”

Kirchhoff’s Law states that the emissivity of a material is equal to the absorptivity of that material on a wavelength-by-wavelength basis. If the emissivity is low over a spectral range, then the absorptance will also be low over that range. The reflectance is generally high for such surfaces, especially if the coating and/or glass substrate are not very transmissive over the spectral range involved. This is because the sum of the absorptance, transmittance, and reflectance of a surface is 1.0. If two of these properties are low, the other must be high and vice versa. Thus the “low-e” designation applied to window glass refers to the emissivity (low) and reflectance (high) properties of the coated surface over the long-wavelength FIR spectral region of surface-emitted radiation.

The emissivity of a glass surface over the solar portion of the spectrum is not likely to be low because that implies high absorptance and a very dark window, one that would not transmit substantial quantities of solar or visible radiation. A low-e coating placed on surface 3 works by blocking the FIR radiative heat transfer, reducing its emission from the warmer inner pane. When it is placed on surface 2, it takes advantage of the coating’s high reflectivity over the FIR, which causes much of the radiated heat from the warm inner pane (surface 3) to be reflected back to that pane from the colder outer pane (surface 2) before the radiant heat can be absorbed by the outer pane and then conducted and convected outdoors as a heat loss from the interior.

It is a quirk of common practice that radiative issues are embedded in a discussion of the conductive properties of a glazing system. The U-factor is but an approximation, a simplification of the more general heat transfer properties of a glazing system for the purposes of simplifying engineering calculations of window energy performance.

The second kind of coating commonly used in IGUs is a modification of the first. In this case, the purpose is to adapt the coating to the needs of buildings with significant cooling periods—when daytime solar radiant heat gain is as important as (or more so than) conductive heat loss on cold winter nights. In this case, the high spectral reflectance characterizing the FIR region for high solar-gain low-e coatings is extended toward lower-wavelength radiation across the near-IR region to the edge of the visible portion of the spectrum, at which point this reflectivity ideally should drop to zero and permit good transmission of visible daylight, the primary purpose of the window.

This coating is called a low solar gain low-e coating because it reflects near IR solar radiation back outdoors before it can be absorbed in the glazing or transmitted directly to the interior. Invisible solar IR radiation only enters a building as heat and is therefore unwanted in hot climate regions. This relatively new glass coating helps

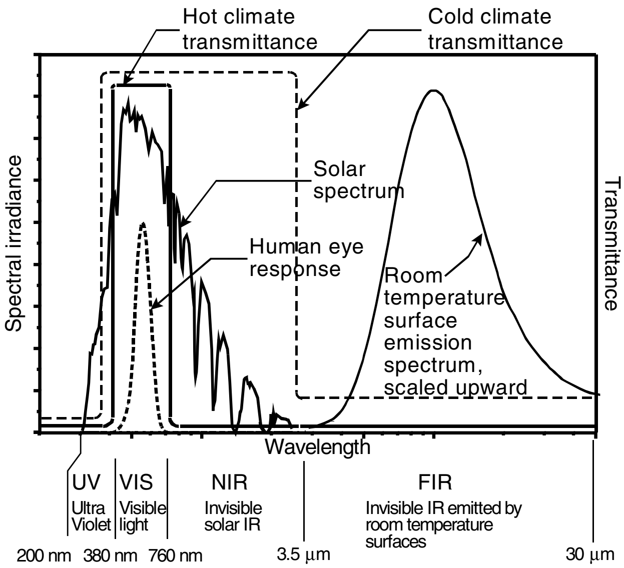


Fig. 7 Spectral irradiances of solar radiation and the infrared emission irradiance from a blackbody at room temperature. The blackbody curve is scaled so that its maximum value is the same as the solar spectrum maximum. Also shown are idealized spectral transmittances of both high and low solar gain low-e coatings.

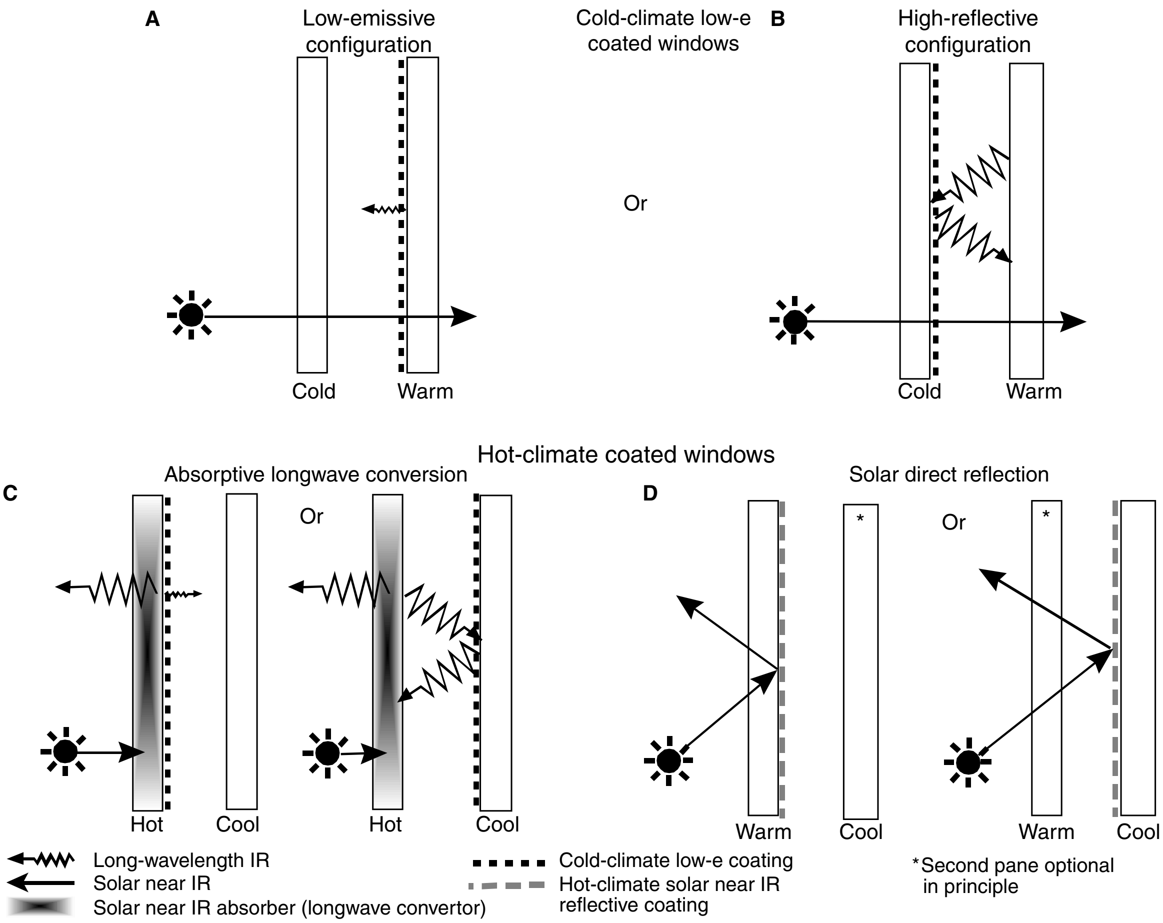


Fig. 8 Coatings and tints for energy efficient glazing systems.

keep a building's interior cool on hot summer days without placing an unnecessary heat burden on the air cooling system, while still admitting ample quantities of natural daylight. The idealized spectral transmittances of the two types of coatings are illustrated in Fig. 7.

Fig. 8 illustrates several ways these two coatings and spectrally selective absorption in the glass can be used to accomplish energy and illumination objectives in both hot and cold climates.

In Fig. 8, A and B describe the two cold-climate coated window configurations. A has the high-solar-gain low-e coating on surface 3 to reduce the wintertime emission of heat from the warmer inner pane to the colder outer one while B has the coating on surface 2 to reflect the long wavelength IR radiation emitted by the warmer inner pane toward the colder outer one. Likewise, parts C and D illustrate the operation of low-solar-gain low-e coatings in reducing summertime solar heat gain through double pane windows. C illustrates the absorptive case in which solar radiation is absorbed within the outer glass and the emission of the absorbed heat toward the colder inner pane is suppressed by a high solar gain type low-e coating on surface 2 by the low emittance of that coating or it is suppressed by reflection from the low-e coating on surface 3, sending the reflected heat back to the outer pane which dissipates it to the outside air. D illustrates the reflective case in which solar radiation is not absorbed by the glass but rather is reflected from a special low-solar-gain low-e coating on either surface 2 or surface 3, sending that radiation back outdoors, without much of it being absorbed within the window glass. The coating in this case, though reflective to the infrared portion of the solar spectrum, is transmissive over the visible portion, thereby allowing good visible transmittance for good view and daylight admission. In some cases, it is less expensive to increase the chemicals responsible for NIR absorption in glass than to make a low solar gain low-e coating. The high solar gain low-e coating generally has fewer layers and—for this and other reasons—is therefore less expensive to manufacture.

CONDENSATION RESISTANCE

Water vapor generally condenses from the vapor phase to liquid droplets when the temperature of the air or a surface in contact with the air drops below what is known as the dew point temperature, which is the temperature below which water vapor mixed with air at a specific humidity level can no longer remain in the vapor phase. The greater the air's humidity, the higher the dew point temperature. Condensation on the surfaces of a window is common with insufficiently insulated windows and most often seen when it is cold outside and warm inside. In buildings well-sealed against cold winter air and filled with relatively warm air, the moisture content indoors tends to rise due to several interior sources of water vapor. If a window is not very well-insulated, the window glass and/or frame can lose heat to the cold outside, dropping their interior surface temperatures. When these temperatures drop below the dew point for the indoor air near the window, condensation forms. Condensation can also occur on air conditioned buildings during humid weather conditions, with the condensation in this case forming on the outside surfaces of the window.

Condensation is annoying, impedes the view, and can be destructive to the window and building materials. The solution is to insulate the window better, keeping surfaces adjacent to the humid air warmer than the dew point temperature of that air.

The Condensation Resistance Factor (CRF) is a measure of the effectiveness of a window or glazing system in reducing the potential for condensation. The higher the CRF, the more energy efficient the window and glazing system, generally. Thus, a low U-factor generally leads to a high CRF, as long as there is not a small "thermal short-circuit" in the window system that allows an interior surface to become too cool.

THE "SOLAR HEAT GAIN COEFFICIENT"—SOLAR RADIANT HEAT GAIN

The solar heat gain property of a fenestration is the sum of two factors. See Fig. 5. The first is characterized by the solar transmittance of the window, T_s, given by

$$T_S = \frac{\int T(\lambda) E_{S\lambda}(\lambda)\, d\lambda}{\int E_{S\lambda}(\lambda)\, d\lambda} \tag{3}$$

where $E_{S\lambda}(\lambda)$ is the solar spectral irradiance distribution and $T(\lambda)$ is the spectral transmittance. The second is the inward-flowing fraction N_i of the window system's solar absorptance A_s, given by

$$N_i A_S = N_i \frac{\int A(\lambda) E_{S\lambda}(\lambda)\, d\lambda}{\int E_{S\lambda}(\lambda)\, d\lambda} \tag{4}$$

where $A(\lambda)$ is the spectral absorptance. Added together, the two of these form what is called the Solar Heat Gain Coefficient (SHGC).

$$\text{SHGC} = T_s + N_i A_s \tag{5}$$

This is the ratio of total solar heat gain through the window to the incident solar irradiance on the window. The solar heat gain Q_s through the window is the product of the incident solar irradiance E_s over the plane of the window, the area A of the window, and the SHGC value for the window.

$$Q_s = E_s A \ \text{SHGC} \ [\text{watts}] \tag{6}$$

VISIBLE TRANSMITTANCE, UV TRANSMITTANCE, AND DAMAGE-WEIGHTED TRANSMITTANCE

The human eye can see only wavelengths over the approximate range from 380 to 760 nm. Radiation outside this range is not visible and should not be called light. The eye's spectral response follows a bell-shaped curve, seen in Fig. 1 of the entry on daylighting, below the solar spectrum therein plotted, peaking at 555 nm. The visible transmittance T_v of a glazing system is the ratio of luminous flux transmitted by the window to the luminous flux incident upon it, defined by the following equation:

$$T_V = \frac{\int T(\lambda) V(\lambda)\, d\lambda}{\int V(\lambda)\, d\lambda} \tag{7}$$

where $V(\lambda)$ is the photopic spectral luminous efficiency function or "V-lambda" weighting function and $T(\lambda)$ is the spectral transmittance of the glazing system. T_V is also known as VT.

The ultraviolet transmittance of a glazing system is also of interest because many people associate the transmission of ultraviolet radiation with fabric fading and other damage to a building's interior. There have been studies of fading and other damaging effects of solar radiation, but no consensus has yet emerged on which portion of the solar spectrum is most responsible or what spectral weighting function is appropriate for assessing the contribution of different solar UV wavelengths to the damage in a single "UV transmittance" figure. For a while, the straight integrated spectral transmittance from 300 to 380 nm was used, given the symbol T_{-UV}. More recently, the spectrum of interest has been extended beyond the UV portion to cover the range from 300 to 700 nm, and a different weighting function was selected. With this system, the $V(\lambda)$ weighting function of Eq. 7 is replaced by another function purported to represent the damaging portions of the solar spectrum. The damage-weighting function is

$$S_{dm,rel}(\lambda) = e^{3.6 - 12.0\lambda} \lambda \ \text{in} \ \mu\text{m} \tag{8}$$

The resulting damage-weighted transmittance has the symbol T_{-dw}. The methodology is based on the work of

Win–Wire

Jurgen Krochmann in Germany and stems from his studies of the damaging effects of radiation on paintings and other museum artifacts. The Krochmann damage-weighting function was incorporated into ISO/CIE publication 89/3 "On the Deterioration of Exhibited Museum Objects by Optical Radiation" and is referenced by standard NFRC 300 optical properties standard in computing the damage-weighted transmittance $T_{-\mathrm{dw}}$.[3]

ILLUMINATION

Daylight illumination from windows and other fenestrations can be used to displace electric lighting energy use. The conditions for this to be successful are that: (1) the space should be occupied by people (in need of illumination); (2) the building should be designed for ample, high quality illumination from the fenestration systems; (3) the electric lighting should be turned off or dimmed to save energy when daylighting is adequate; and (4) the sun and sky conditions are such that the daylight illumination incident on the building is adequate. Design for daylighting is discussed in the entry on daylighting.

ENERGY PERFORMANCE

Solar radiant heat gain through windows, which is generally desirable during the winter and undesirable in the summer, and conductive heat transfer are the principal sources of unwanted energy costs associated with fenestrations. They can add to the cost of heating the building in winter and cooling it in summer. Computing the energy costs and savings associated with a building's fenestration systems has been eased by a variety of computer simulation tools that are now available. The WINDOW program, coupled with THERM for calculating 2-D frame and edge effects, both developed at Lawrence Berkeley National Laboratory, allows rapid calculation of the instantaneous energy performance characteristics of a window.[4] The results of these calculations can be used in computerized hourly building energy performance computer programs. RESFEN is another LBNL computer program for calculating the heating and cooling energy use of windows in residential buildings over a long time period. All of these LBNL computer programs are available for free download at http://windows.lbl.gov/software/. Additional energy simulation tools are available for determining fenestration contributions to a building's energy performance. They are described at the LBNL web site and some are available there. Additional tools can be found via a Web search.

WINDOW ECONOMICS

Replacing old windows in a building with energy-efficient ones and installing high performance windows during new construction can result in significant yearly energy savings, increased comfort, and improved occupant satisfaction. While these are desirable in residential buildings, better windows can also increase productivity in commercial or office buildings. Occasionally, the productivity enhancements save more dollars than the entire energy bill itself.

Achieving these desirable results can be expensive. (There are exceptions, such as the case where reduced solar heat gain from improved windows allows the air conditioning equipment to be significantly smaller, enough so that the saved equipment costs offset the extra window costs.) In markets driven primarily by least-first-cost goals, it can be difficult to justify the extra costs of better windows. Several accounting tools are available for calculating the longer-term benefits. These include payback time, return on investment, cash flow analysis, life-cycle cost/benefit analysis, and net energy analysis. Using these tools, the designer can capture at least some of the future dollar savings attributable to better windows. But these measures generally fail to include a variety of additional benefits, such as the dollar savings of improved worker productivity, the economic benefits of a healthier society, the values of improved ecosystem health, or the influence of future energy prospects.

CONCLUSION

As we pass the peak of world oil production, entering a subsequent steady decline in oil's availability, and as we approach future peaks in the production of coal and natural gas, the rising price of fossil fuel energy will shorten payback times and improve other indicators of window economic performance. Government actions, such as green taxes and other positive and negative incentives, can speed the adoption of energy-efficient and otherwise high-performance fenestration systems. Large projects offer higher budgets for design and specification, simplifying decisions to incorporate some of these additional benefits in design and product selection. For smaller projects, perhaps the best advice would be to use the most energy-efficient and affordable window system for comfort and as a hedge against future energy price increases.

ACKNOWLEDGMENTS

Portions of the Glossary section have been excerpted, with permission, from "Photometry" by Ross McCluney, in

Encyclopedia of Optical Engineering, Marcel Dekker, New York, 2003.

REFERENCES

1. International Lighting Vocabulary, CIE Publ No. 17.4 IEC Publ. 50(845), Central Bureau of the Commission Internationale de l'Eclairage: Kegelgasse 27, A-1030 Vienna, P.O. Box 169, Austria.
2. McCluney, R. *Introduction to Radiometry and Photometry*; Artech House: Boston, MA, 1994.
3. National Fenestration Rating Council. 8484 Georgia Avenue, Suite 320, Silver Spring, MD 20910, http://www.nfrc.org.
4. Building Technologies Dept., Building 90 Room 3111, Lawrence Berkeley National Laboratory: 1 Cyclotron Road, Berkeley, CA, 94720.

Window Films: Savings from IPMVP Options C and D☆

Steve DeBusk
Energy Solutions Manager, CPFilms, Inc., Martinsville, Virginia, U.S.A.

Abstract
Building owners and property managers have provided numerous reports of the energy savings obtained from the retrofit installation of energy-control window film to existing buildings. In many cases, substantial energy savings are reported with simple paybacks often within one to four years. However, such claims have not been produced using the rigorous methods prescribed by the International Performance Measurement and Verification Protocol (IPMVP, see www.ipmvp.org).

To provide for more definitive evidence of the savings from the use of energy-control window films, CPFilms undertook a demonstration project to install window film on a mid-sized commercial office building and then measure the resulting energy savings using two separate IPMVP methods. The measurement of savings involved the use of IPMVP Option C (whole-building method using utility bill analysis) and Option D (calibrated simulation). During the twelve months following film installation, Option C measured savings of 8.8% were achieved, which includes savings over both the cooling and heating seasons. This compares favorably to Option D calculated savings of 8.4% that were estimated prior to film installation. Both estimates of energy savings demonstrate a simple payback in less than three years.

PROJECT OVERVIEW

In early 2002, CPFilms completed the installation of 9200 ft^2 of energy-control window film on an eight-story building in Rockford, Illinois (near Chicago). Fig. 1 shows the building chosen for the demonstration project following completion of window film installation. The building's conditioned space is 59,000 ft^2 and the building's windows are single-pane bronze tinted glass. The building is heated and cooled using all-electric room unit ventilators.

The window film chosen for this demonstration project was CPFilms' LLumar® E-1220 Low-E film. As shown in Table 1, addition of this film reduced solar heat gain 67% based on the solar heat gain coefficients (SHGC) of the windows before and after film installation. In addition, the film is constructed with a low-emissivity, or low-e coating that improved the insulating value (U-value) of the single-pane windows by 23%. This improvement in the window's insulating properties provided additional cooling-season energy savings, beyond the reduction in solar heat-gain, and provided heating-season energy savings by reducing heat loss through the windows.

Rockford has over 6000 heating degree days per year (using a 65°F base). This cool-climate location was chosen to demonstrate that, under the proper circumstances, energy-control window films are an attractive option even in cooler climates and are not simply a "warm climate only" technology, especially when utilizing low-e insulating films. The location was also chosen to show that low-e window films, and the improvement in window insulating properties they offer, can provide substantial energy savings during a prolonged heating season, more than offsetting the loss of "free" solar heat. Negotiations with the building owner prohibited other energy-saving measures from taking place in order to isolate the effects from window film installation.

Before film installation, CPFilms utilized DOE-2.1 to evaluate project energy savings using IPMVP Option D calibrated simulation methods. CPFilms then enlisted the efforts of Johnson Controls, a leading national energy service company (ESCO), to independently measure and verify energy savings using IPMVP Option C. The Option C measurement of energy savings was compared to predicted savings from the calibrated simulation as a means of determining the suitability of using DOE-2 in predicting energy savings for future projects. This was deemed important, as normally it is not possible to isolate the energy savings from window film installation. Often multiple energy-saving measures are performed along with window film. As a result, Option D typically is the only viable means for determining energy savings from window film installation.

WINDOW FILM ENERGY SAVINGS USING IPMVP OPTION D

Prior to window film installation, a whole-building energy simulation of the demonstration building was prepared.

☆ This entry originally appeared as "Measuring Savings from Energy-Control Window Films Installations Using IPMVP Options C and D" in *Energy Engineering*, Vol. 102, No. 4, 2005. Reprinted with permission from AEE/Fairmont Press.

Keywords: Building envelope; DOE-2; Energy-control films; IPMVP; Low-emissivity; Solar heat gain coefficient; Windows; Window film.

Fig. 1 Demonstration building following film installation.

Table 2 Values used in doe-2 simulation

Building type	Commercial office
Building size	8 floors, 59,000 ft^2 total conditioned area
Window areas	North—2875 ft^2
	South—2875 ft^2
	East—1725 ft^2
	West—1725 ft^2
Window type	Single-pane, 1/4-in., Bronze
Lighting offices	1.0 w/ ft^2
Plug load	w/ ft^2
Heating temperatures	70°F daytime, 65°F night
Cooling temperatures	75°F daytime, 85°F night
Ventilation rate	20 cfm per person, 250 ft^2 per person
Cooling system efficiency	1 kW/ton

To ensure that the simulation model would provide an accurate assessment of the energy savings resulting from window film installation, the monthly and annual energy consumption of the simulated building was compared to the building's actual consumption for the two previous years. DOE-2 energy analysis software was chosen as the method for simulating the building's energy performance, based upon the regular use of the program in the research, ESCO, and utility industries as a reliable method for such purposes. Table 2 provides a review of the most pertinent information used in the simulation.

Fig. 2 shows the comparison of monthly and annual electricity consumption for the building using averaged data from 2000 to 2001 compared to the simulation results. As indicated in Fig. 2, the simulation estimate of annual energy use for the demonstration building closely matches actual data for the 2000–2001 period, while matching the monthly fluctuations in energy use reasonably well. As the simulation model uses averaged weather data over multiple years, it is not reasonable to expect simulation and actual monthly energy use to match precisely each month.

Table 1 Window thermal and solar properties

Window values	U-value[a] (Btu/ h/ ft^2/deg F)	Solar heat gain coefficient
Before film installation	1.09	0.64
After film installation	0.84	0.21
% Improvement with film	23%	67%

[a]National Fenestration Rating Council (NFRC)/American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Winter night-time conditions, 0°F outdoors, 15 mph wind, 70°F indoors.

Next, the building DOE-2 model was modified to show the effect of adding CPFilms' LLumar® E-1220 low-e window film to the existing windows. As seen in Table 3, the DOE-2 calculations indicated an 8.4% reduction in annual electricity consumption, saving approximately 180,850 kWh annually and producing a simple payback in 2.77 years.

WINDOW FILM ENERGY SAVINGS USING IPMVP OPTION C

Energy savings for the demonstration building were also independently measured by Johnson Controls using IPMVP Option C, employing the analysis of monthly utility bills. This method was deemed appropriate, as the demonstration building is all electric, making it relatively simple to identify the energy savings from window film installation for both the heating and cooling seasons.

As expected, the main factors causing fluctuations in energy consumption for the building, based on a review of prior energy use, were: (1) building occupancy, and (2) weather conditions. As mentioned, other factors affecting building energy consumption (type and amount of lighting, heating, ventilating and air-conditioning (HVAC) equipment efficiency and type, cooling and heating temperatures, etc.) were held relatively stable to enable isolating the savings from window film installation as much as possible. Accordingly, the method for measuring

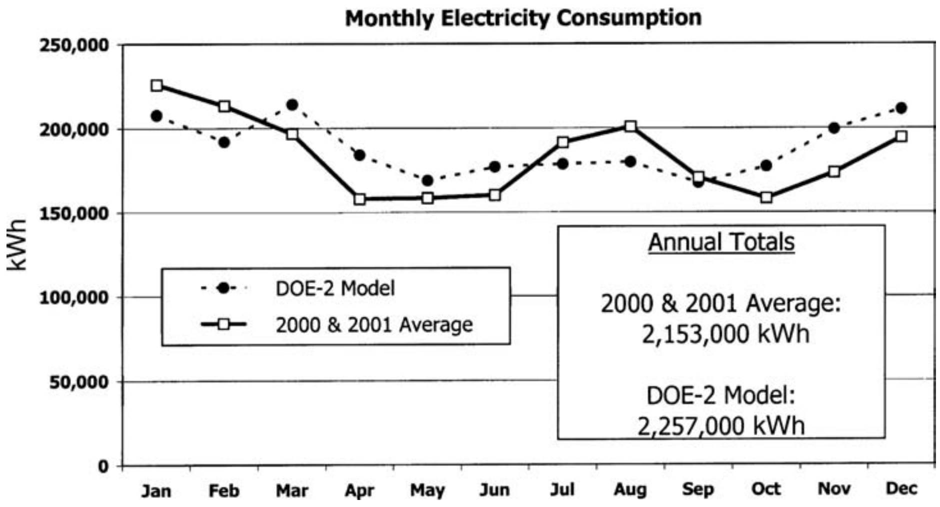


Fig. 2 Monthly and annual electricity consumption (actual vs DOE-2 model).

Table 3 Annual savings using Option D

Existing annual electric consumption and cost based on average of 2000 & 2001	2,153,000 kWh	$138,222
Estimated savings (as a % of existing consumption)	8.4%	
Estimated annual savings	180,850 kWh	$11,610
Simple payback	2.77 years	

energy savings would be to determine the energy consumption over the 12-month period prior to window film installation, after adjusting consumption for occupancy and weather, and compare this figure to the value for the 12 months following window film installation.

Fig. 3 shows the electricity use per thousand ft^2 of occupied space per degree day (kWh/ksqft/DD), before, during and after window film installation. From Fig. 3 it is apparent that during and immediately following window film installation, energy usage began to decrease significantly. As seen in Fig. 4, the average energy consumption for the twelve months prior to film installation was 11.18 kWh/ksqft/DD. Following film installation, the same information was tracked on a monthly basis and the average energy consumption for the twelve months following film installation was 10.19 kWh/ksqft/DD. A comparison of these "before" and "after" values indicated the project's overall savings to be 8.8%.

Table 4 shows the results obtained from both IPMVP methods, in a side-by-side comparison. As indicated in this table, the results of both methods match closely, giving additional confidence to the results obtained.

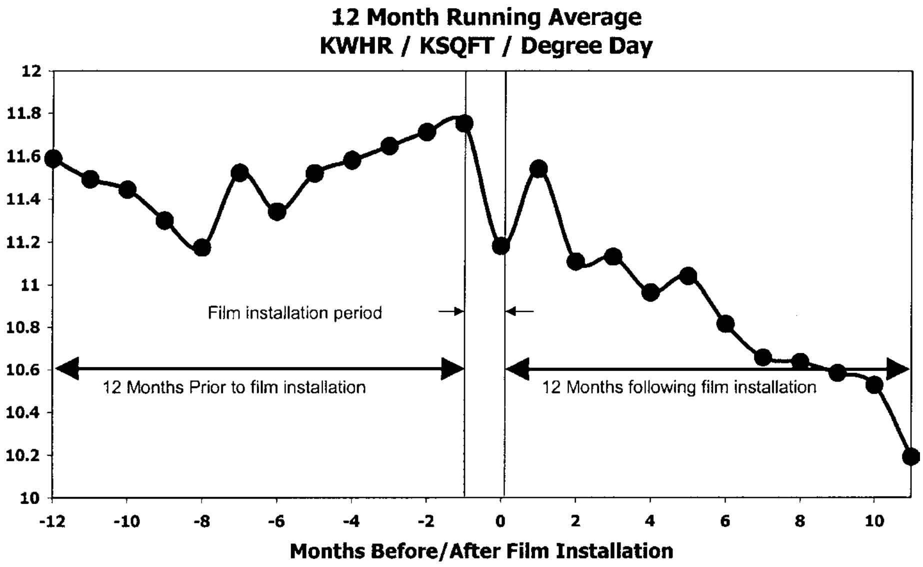


Fig. 3 Building energy use—before and after film installation.

Win–Wire

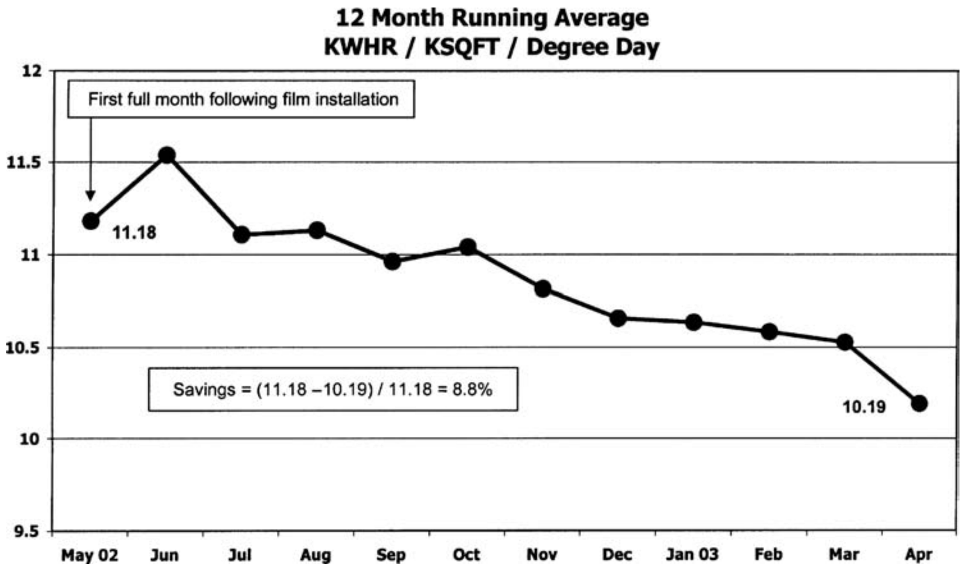


Fig. 4 Determining energy savings after film installation.

Table 4 Comparison of energy savings using Option C and Option D

	Option C	Option D
Annual kWh savings	189,460	180,850
Savings as a % of existing consumption	8.8%	8.4%
Annual dollar savings	$12,163	$11,610
Simple payback (years)	2.65	2.77

CONCLUSION

The demonstration project provided insight into the substantial savings that can be achieved from the installation of energy-control window films. The project also demonstrated that significant energy savings can be obtained even in cool climates, if a significant amount of energy-inefficient window glass is utilized on the building envelope. Energy savings in warmer regions are likely to be greater than those shown in this article, but actual savings will be dependent upon many variables, especially existing glass type and type of window film utilized. These variations can easily be accounted for within a DOE-2 analysis for an accurate assessment.

The project also demonstrated that energy savings from window film installation can be accurately predicted using DOE-2 or other appropriate computer simulation software, if an accurate model of the building is prepared. This is especially important, as directly measuring the energy savings from window film installation using IPMVP methods other than Option D is often difficult, as the many factors affecting energy consumption are not normally held constant as they were in this study. Using the information provided by this project, energy savings from future window film projects can be predicted with confidence using DOE-2.

ACKNOWLEDGMENTS

The author wishes to thank the following for their assistance: Jeff Hongsermeier, owner of Glass Enhancements, Rockford, Illinois, for providing window film installation services and for his assistance in developing this project, Johnson Controls for assistance in project development and for determining Option C savings, and the Fridh Corporation for use of their property.

Window Films: Solar-Control and Insulating

Steve DeBusk
Energy Solutions Manager, CPFilms Inc., Martinsville, Virginia, U.S.A.

Abstract
Solar-control films are used worldwide as an effective means of lowering building energy costs by reducing excessive solar heat gain through windows. Determining energy savings from actual installations is difficult due to the common practice of implementing several energy conservation measures simultaneously and due to annual variations in the many factors that can affect a building's energy usage.

To isolate and quantify the energy-saving benefits of solar films, an energy analysis study on a conventional office building was undertaken using the U.S. Department of Energy (DOE)'s sophisticated DOE-2 energy-simulation software. The study included several types of solar films applied to various glazing systems. Various cities were included in the study to illustrate the energy savings in different climates and to show the effect of differing electricity costs.

Based on the most typical types of installations and on customary installation costs for medium-size commercial projects, the average return on investment (payback) from solar film application was an impressive 2.65 year. These savings were the result of reducing annual electricity kilowatt-hour (kWh) usage by an average of 6.6% and reducing peak summer month kilowatt demand on average by 6.4%.

INTRODUCTION

Solar-control window films are considered in the building industry to be "retrofit" products—that is, products that are applied to existing buildings after construction as opposed to being used in new construction.[1,2]

Solar Film Construction

Solar-control window films typically consist of a thin (0.025-mm, 0.001-in.) polyester film substrate that has a microthin transparent metal coating applied to one side. This metal coating is applied using vacuum-based technologies such as vapor deposition or sputtering. The metal coating may be a single metal, an alloy, a metal-oxide, or a combination of these coatings. A second layer of polyester film is laminated over the metal coating to protect the metal. Onto one side of this laminated composite, an acrylic scratch-resistant (SR) coating is applied to the surface that will face the building interior. This SR coating protects the film during normal window cleaning. On the opposite side of this film laminate, a clear adhesive is applied, which will eventually bond the film to the window glass. This adhesive layer is protected by a removable release liner until just before field application. The film is protected from ultraviolet (UV) degradation by UV absorbers that are added to the polyester film layers, the adhesive layer, or both.

Keywords: Building envelope; DOE-2; Energy simulation; Solar control; Solar films; Window films; Windows; Window systems.

Solar Film Appearance and Properties

The appearance of the film (the color, the level of visible light transmission, and the degree of reflectivity) depends on the metal coatings used. Typical all-metal solar films can be silver-reflective, gray, silver-gray, bronze, or light green in color. Visible light transmissions can vary from very dark (10%) to very light (70%), and visible reflectance can vary from the same reflectance as clear glass (8%) to highly reflective (60%). The ability of a glazing system to reduce solar heat gain is measured by its solar heat-gain coefficient (SHGC). As expected from the variety of films available, the SHGC for solar films can vary significantly, from 0.14 to 0.69, as measured on 6-mm (1/4-in.) clear glass.

Solar Film Benefits

This combination of film properties produces a product that provides several important benefits:

- Reduced cooling energy costs by reducing excessive solar heat gain
- Enhanced reduction in cooling and heating energy costs when low-e type films are used
- Enhanced tenant comfort from improved temperature distribution (less hot and cold spots) and reduced glare
- Uniform building appearance from the exterior—improving tenant retention in leased buildings
- Reduced fading of carpets, drapes, and furnishings due to the UV-blocking ability of films
- Privacy for building occupants when using reflective or dark films

Solar Film Installation Process

The first step in the installation of solar film involves rigorous cleaning of the window glass surface. Next, an application solution is sprayed onto the glass; the release liner is removed from the film; and the adhesive side of the film is carefully placed onto the glass surface. The application solution allows one to move the film on the glass for precise film placement, and it prevents air from becoming trapped between the film and glass. Then the film is carefully trimmed around the perimeter of the window, leaving a 1–3-mm (1/32-in. to 1/8-in.) gap between the film and frame. Water is sprayed onto the film surface, and a rubber squeegee is used to remove the application solution and bond the film to the glass surface.

MOTIVATION FOR SOLAR-CONTROL FILM ENERGY ANALYSIS STUDY

Solar-control films have been used since the early 1960s as an effective means of reducing building energy costs. Unfortunately, it is difficult to determine precisely the cost savings from applying film to a building, due to variations in the many factors that affect a building's energy consumption from year to year (weather changes such as the amount of sunshine and temperature/wind differences, changes in occupancy, upgrades or other changes in building energy-using equipment, changes in maintenance for key equipment, the addition of energy-consuming equipment such as computers, etc). This situation is usually complicated by the fact that building owners usually perform energy-conservation upgrades such as solar-control film application in conjunction with other upgrades, making it impossible to determine the savings from film application alone. Therefore, a means of accurately estimating the energy savings from solar film application was needed.

One of the most accurate and reliable energy simulation software packages available is the DOE-2 energy analysis program.[3] DOE-2, which uses an hourly calculation method, has been validated many times by comparing its results with thermal and energy-use measurements on actual buildings (see http://gundog.lbl.gov).[4] DOE-2 is used worldwide by energy engineers, architects, government organizations, and utilities as a means of estimating the effect of various measures on a building's energy consumption and for developing building energy codes. As a result, it was determined that a reasonable course of action was to use DOE-2 modeling to estimate the energy savings from application of solar-control window film.

SCOPE OF ENERGY ANALYSIS STUDY

The DOE-2 energy study was performed on a conventional (1990s) 10-story office building with a total floorspace of 16,257 m^2 (175,000 ft^2). To gauge the effect of different films, four films were chosen and categorized as "Maximum-Performance," "Maximum-Performance

Table 1 Solar and thermal performance factors for windows and films in study

	No film	Maximum performance film	Maximum performance low-E film	High performance film	High performance low-E film
Solar heat-gain coefficient (SHGC)					
Single clear	0.81	0.23	0.17	0.36	0.28
Dual clear	0.70	0.31	0.27	0.42	0.35
Single gray	0.57	0.27	0.21	0.33	0.26
Dual gray	0.45	0.24	0.21	0.30	0.25
U-values (W/m^2 °C)					
Single clear	6.18	5.81	4.78	5.87	4.69
Dual clear	2.74	2.65	2.35	2.66	2.32
Single gray	6.19	5.81	4.78	5.87	4.69
Dual gray	2.74	2.65	2.35	2.66	2.32
U-values (BTU/h/ft^2/°F)					
Single clear	1.09	1.02	0.84	1.03	0.83
Dual clear	0.48	0.47	0.41	0.47	0.41
Single gray	1.09	1.02	0.84	1.03	0.83
Dual gray	0.48	0.47	0.41	0.47	0.41

Low-E," "High-Performance," and "High-Performance Low-E," based on the film's SHGC on single-pane, 6-mm (1/4-in.) clear glass.

The study was performed on single-pane clear, dual-pane clear, single-pane gray-tinted, and dual-pane gray-tinted window systems. Each film type was analyzed on each of these glazing systems. All windows consisted of 6-mm-thick (1/4-in. -thick) panes, and for the dual-pane units, the panes were separated by a 12-mm (1/2-in.) air space.

The SHGC and Winter Night-Time *U*-values for all glazing systems contained in the study are shown below. The SHGC shown is at normal solar incidence angle, and the Winter Night-Time *U*-value is based on American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) standard Winter conditions, using a −17.8°C (0°F) outdoor air temperature, an indoor air temperature of 21°C (70°F), a 24.1 km/h (15 mph) wind speed outdoors, a 0 km/h (0 mph) wind speed indoors, and 0 W/m^2 (0 Btu/h/ft^2) solar intensity. In the analyses, the SHGC was calculated (by DOE-2) for the precise sun angle at each hour of the day; and the window *U*-value was calculated at each hour based on indoor and outdoor temperatures, outdoor air speed, and solar intensity (Table 1).

The model building was a square building with equal glass area facing north, south, east, and west. The glass area on each of the four building exposures was 557.4 m^2 (6,000 ft^2). Models with film applied used film on the east, south, and west exposures only.

Other model parameters, typical of modern office buildings, used in the study included

- Indoor lighting 10.76 W/m^2 (1.0 W/ft^2)
- Office equipment 11.95 W/m^2 (1.1 W/ft^2)
- Heating setpoint 21.1°C (70°F)
- Heating setback 18.3°C (65°F)
- Cooling setpoint 23.9°C (75°F)
- Cooling setback 26.7°C (80°F)
- Medium-colored blinds used 25% of the time; SHGC of blinds 0.69.
- Windows recessed from building face 15 cm (6 in.) providing partial shading of all windows
- Variable-air-volume (VAV) air-distribution system and air-side economizer
- Heating plant using gas boilers with an efficiency of 80%
- Chillers with full-load efficiency of 0.69 KW per ton (Coefficient of Performance [COP] of 5.1)

The parameters used (such as the use of blinds, recessed windows, the VAV air distribution system, the air-side economizer, and the high-efficiency chiller system) effectively reduce the savings from solar film application. This was desired to provide reasonable and conservative estimates of energy savings from solar film installation, not to create a "best case" scenario.

Electricity costs for each location were determined from the commercial rate schedules published on the Web sites for electric utilities in each city. Rate schedules that applied to buildings with peak kilowatt demands of approximately 1000 KW were used. Both kilowatt-hour and kilowatt demand charges were used. The rate schedules used typically vary the kilowatt-hour and kilowatt charges by time of year and time of day, and are too complex to provide here; however, to provide the reader a general idea of the costs used, Table 2 shows the average costs for each city based on the total annual kilowatt-hour used and the total annual electricity costs for the four building models without film.

RESULTS OF STUDY

Tables 3–5 show the payback, reduction in annual kilowatt-hour usage, and reduction in summer month

Table 2 Average electricity costs by location

City	Average cost ($/kWh)	Electric utility and rate schedule
Boston	0.1220	Boston edison, rate G-3
Chicago	0.0892	Commonwealth edison, rate 6L
Dallas	0.0907	Texas-new mexico power, large general service
Jacksonville	0.0697	Florida power & light, GSLD-1
Los Angeles	0.1336	Southern California edison, TOU-8, large general service
Memphis	0.0604	Memphis light, gas & water, general service GSA
Phoenix	0.0595	Arizona public service co., general service
Toronto (Canada)	0.0553	Toronto hydro-electric, business rates
Washington, D.C.	0.0706	Baltimore gas & electric, schedule GL (option 2)
Overall average	0.0834	

Natural gas was used as the heating fuel and for domestic hot water production at a cost of 70 cents per therm for all locations.

Table 3 Simple payback by location, window, and film type

	Boston	Chicago	Dallas	Jacksonville	Los Angeles	Memphis	Phoenix	Toronto	Washington	Average
Single clear										
Max-perf film (%)	0.89	1.11	0.87	1.18	0.69	1.42	1.10	1.73	1.28	1.14
Max-perf low-E film (%)	0.81	1.00	0.85	1.15	0.69	1.33	1.07	1.43	1.16	1.06
High-perf film (%)	1.12	1.37	1.10	1.53	0.87	1.82	1.38	2.19	1.66	1.45
High-perf low-E film (%)	0.91	1.10	0.95	1.31	0.79	1.48	1.20	1.55	1.30	1.18
Single clear-all film types average (%):										1.21
Dual clear										
Max-perf film (%)	1.46	1.79	1.42	2.03	1.12	2.57	1.77	2.64	2.12	1.88
Max-perf low-E film (%)	1.34	1.65	1.34	1.90	1.08	2.29	1.70	2.29	1.93	1.72
High-perf film (%)	2.01	2.50	2.00	3.02	1.57	3.69	2.47	3.94	3.14	2.70
High-perf low-E film (%)	1.55	1.91	1.60	2.31	1.25	2.68	2.01	2.69	2.30	2.03
Dual clear-all film types average (%):										2.09
Single gray										
Max-perf film (%)	2.09	3.14	2.18	2.81	1.78	3.81	2.39	4.16	2.90	2.81
Max-perf low-E film (%)	1.42	1.94	1.59	2.09	1.36	2.49	1.90	2.34	1.96	1.90
High-perf film (%)	2.62	3.95	2.81	3.70	2.26	4.96	2.99	5.31	3.84	3.61
High-perf low-E film (%)	1.49	1.98	1.73	2.28	1.49	2.61	2.03	2.39	2.08	2.01
Single gray-all film types average (%):										2.58
Dual gray										
Max-perf film (%)	3.92	5.57	3.71	4.63	2.99	6.90	3.86	6.53	4.96	4.79
Max-perf low-E film (%)	2.90	3.56	2.81	3.57	2.32	4.65	3.18	4.26	3.57	3.42
High-perf film (%)	5.60	8.02	5.49	6.82	4.34	10.26	5.20	9.19	7.04	6.88
High-perf low-E film (%)	3.12	3.86	3.19	4.04	2.62	5.11	3.58	4.53	3.99	3.78
Dual gray-all film types average (%):										4.72
Location average (%)	2.08	2.78	2.10	2.77	1.70	3.63	2.36	3.57	2.83	
All location, all window types, all film types (%)	2.65									

Table 4 Reduction in annual kilowatt-hour usage

	Boston	Chicago	Dallas	Jacksonville	Los Angeles	Memphis	Phoenix	Toronto	Washington	Average
Single clear										
Max-perf film (%)	9.5	9.6	12.1	11.4	11.7	11.0	13.6	9.2	10.3	10.9
Max-perf low-E film (%)	10.9	10.9	13.5	12.7	12.9	12.3	15.3	10.5	11.7	12.3
High-perf film (%)	7.5	7.8	9.6	8.9	9.3	8.7	10.7	7.2	8.1	8.6
High-perf low-E film (%)	9.4	9.6	11.8	11.0	11.1	10.8	13.4	9.1	10.2	10.7
Single clear-all film types average (%):										10.6
Dual clear										
Max-perf film (%)	6.2	6.4	7.9	7.1	7.6	6.8	9.1	6.2	6.7	7.1
Max-perf low-E film (%)	7.3	7.4	9.2	8.4	8.8	8.0	10.5	7.3	7.9	8.3
High-perf film (%)	4.5	4.6	5.7	4.9	5.5	4.8	6.6	4.3	4.7	5.1
High-perf low-E film (%)	6.1	6.2	7.6	6.9	7.4	6.7	8.8	6.0	6.5	6.9
Dual clear-all film types average (%):										6.8
Single gray										
Max-perf film (%)	4.2	3.8	5.3	5.2	5.0	4.6	6.7	4.0	4.7	4.8
Max-perf low-E film (%)	5.9	5.4	7.3	7.1	6.8	6.5	9.0	5.7	6.5	6.7
High-perf film (%)	3.4	3.0	4.2	3.9	4.0	3.6	5.4	3.1	3.6	3.8
High-perf low-E film (%)	5.4	5.0	6.6	6.3	6.1	5.9	8.2	5.1	5.8	6.1
Single gray-all film types average (%):										5.3
Dual gray										
Max-perf film (%)	2.6	2.5	3.5	3.4	3.3	3.0	4.6	2.8	3.1	3.2
Max-perf low-E film (%)	3.5	3.5	4.7	4.6	4.4	4.1	6.0	3.8	4.2	4.3
High-perf film (%)	1.8	1.8	2.4	2.4	2.3	2.1	3.5	2.0	2.2	2.3
High-perf low-E film (%)	3.1	3.1	4.0	3.9	3.8	3.6	5.2	3.3	3.7	3.8
Dual gray-all film types average (%):										3.4
Location average (%)	5.7	5.7	7.2	6.8	6.9	6.4	8.5	5.6	6.2	
All location, all window types, all film types (%)	6.6									

Table 5 Reduction in summer peak kilowatt demand

	Boston	Chicago	Dallas	Jacksonville	Los Angeles	Memphis	Phoenix	Toronto	Washington	Average
Single clear										
Max-perf film (%)	11.4	10.6	11.2	10.0	9.7	10.1	11.3	11.4	10.4	10.7
Max-perf low-E film (%)	13.0	12.2	12.9	11.5	11.2	11.8	13.6	13.0	12.3	12.4
High-perf film (%)	8.7	8.2	8.5	7.4	7.4	7.6	8.9	8.7	7.7	8.1
High-perf low-E film (%)	10.9	10.4	11.1	9.9	9.1	10.0	11.7	11.1	10.3	10.5
Single clear-all film types average (%):										10.4
Dual clear										
Max-perf film (%)	6.5	6.8	6.9	5.7	5.7	5.8	7.4	7.6	6.3	6.5
Max-perf low-E film (%)	7.8	8.1	8.4	7.1	6.5	7.0	8.9	9.0	7.6	7.8
High-perf film (%)	4.7	4.9	4.8	3.6	4.0	3.7	5.2	5.2	4.1	4.5
High-perf low-E film (%)	6.6	6.6	6.9	5.5	5.8	5.8	7.4	7.3	6.1	6.4
Dual clear-all film types average (%):										6.3
Single gray										
Max-perf film (%)	5.3	4.4	4.9	4.7	4.6	4.4	5.7	5.3	5.4	5.0
Max-perf low-E film (%)	7.4	6.3	7.3	7.0	6.5	6.7	8.3	7.4	7.5	7.2
High-perf film (%)	4.2	3.4	3.8	3.6	3.4	3.3	4.6	4.0	4.2	3.8
High-perf low-E film (%)	6.8	5.9	6.6	6.3	5.4	6.1	7.7	6.6	6.9	6.5
Single gray-all film types average (%):										5.6
Dual gray										
Max-perf film (%)	2.9	2.8	2.9	3.0	2.3	2.6	3.3	3.8	3.2	3.0
Max-perf low-E film (%)	3.7	4.0	4.3	4.2	3.2	3.8	4.7	5.2	4.4	4.2
High-perf film (%)	2.0	1.7	1.9	1.9	1.6	1.7	2.4	2.7	2.4	2.0
High-perf low-E film (%)	3.5	3.4	3.6	3.7	3.0	3.4	4.2	4.5	3.7	3.7
Dual gray-all film types average (%):										3.2
Location average (%)	6.6	6.2	6.6	5.9	5.6	5.9	7.2	7.1	6.4	
All location, all window types, all film types (%)	6.4									

peak demand for each location and for each film and window combination. Also shown are the averages for each window type and each location. The overall average for all locations, window types, and films is also given.

Following are some general observations concerning the results of the study.

- For all window and film types and all locations, the overall average payback for solar film installation was 2.65 year (see Table 3). The average payback by window type: Single Clear, 1.21 year; Dual Clear, 2.09 year; Single Gray, 2.58 year; and Dual Gray, 4.72 year (in almost 50% of the cases involving Dual Gray windows, the payback was less than 4 year).
- As shown in Table 3, it appears that the payback period is affected more by the cost of electricity than by climate effects. The average payback in Boston (2.1 year), for example, is less than the average payback in Jacksonville, Florida (2.8 year), due mainly to the higher average cost of electricity in Boston compared with Jacksonville (12.2 cents per kWh average vs 6.97 cents). Also, the average payback for Memphis (3.6 year) was more than in Washington, D.C. (2.8 year), even though the climate in Memphis is somewhat warmer, solely due to the lower cost of electricity in Memphis.
- The data also show that solar-control film is not a "warm climate only" product. The average payback for cities not considered to be in the Sun Belt was still less than 3 year on average (Boston, 2.1 year; Chicago, 2.8 year; Washington, DC, 2.8 year). The average payback for Toronto (the coolest climate of all cities considered) was still a very respectable 3.6 year, despite the fact that Toronto has the lowest overall electricity prices of cities in the study.
- Solar-control film has a considerable positive effect on reducing annual kilowatt-hour and summer peak demand, on average reducing annual kilowatt-hour usage by 6.6% and summer month peak demand by 6.4% (see Tables 4 and 5).

CONCLUSIONS

This study clearly indicates that solar-control window film can play a useful and viable role in improving the energy efficiency of many buildings and that window films can be effective in reducing energy costs and energy consumption for buildings in many locations. Excellent energy savings can be provided by this technology—typical 5%–10% reductions in peak demand and annual cooling costs with such savings provided within a reasonable payback period (averaging less than 3 year). Although the focus of this entry was locations in the United States, it has been the author's experience (and it should be apparent) that solar-control window films are applicable to a wide range of locations, climates, and countries.

It is important to note that while providing these important energy-saving benefits, window films are able to provide many other benefits that directly hit the mark of key scoring components for "green building" specification programs, such as the Leadership in Energy and Environmental Design (LEED).[5,6] As such, solar-control window films are able to meet the needs of many different design professionals, from property owner/managers to architects to energy engineers to green-building professionals.

REFERENCES

1. International Window Film Association. Advanced Solar Control Guide, 2003.
2. International Window Film Association. Window Film Information Center. Available at www.iwfa.com (accessed on December 2006).
3. Hirsch J.J.and Associates. Welcome to DOE-2.com. Available at www.doe2.com (accessed on December 2006).
4. Lawrence Berkeley National Laboratories. DOE 2.1E. Available at http://gundog.lbl.gov (accessed on December 2006).
5. U.S. Green Building Council. LEED for Existing Buildings Project Checklist. Available at http://www.usgbc.org/DisplayPage.aspx?CMSPageID=221 (accessed on December 2006).
6. U.S. Green Building Council. LEED: Leadership in Energy and Environmental Design. Available at www.usgbc.org/DisplayPage.aspx?CategoryID=19 (accessed on December 2006).

Window Films: Spectrally Selective versus Conventional Applied☆

Marty Watts
V-Kool, Inc., Houston, Texas, U.S.A.

Abstract
Compared to conventional applied window film, high light-transmitting spectrally selective film cost-effectively blocks unwanted solar energy from entering windows and reduces air conditioning costs without darkening building interiors, impeding the ability to see through existing glass, or changing the external appearance of a building.

According to the California Energy Commission, as much as 40% of a building's cooling requirements is a function of heat entering through existing glass. Stopping heat at the window is the most effective means of lowering temperatures and reducing heating, ventilation and air conditioning (HVAC) operating costs. In new construction, reducing heat at the window can mean the need for smaller and less-expensive HVAC systems.

The solution to overheating through windows is to specify solar control glass or applied window film, though even the best solar control glass performs no better than the best applied window film. Solar control glass can be selected for optimum energy performance in reference to the geographic orientation of any given building or section of a building. However, even in new construction, the cost of solar control glass often exceeds the cost of standard glass to which a solar control film is later applied.

For existing buildings experiencing problems from heat through windows, the most expensive option is to replace existing glass and frames with a new window system designed to block heat and deal with a building's energy performance needs. Less expensive is keeping existing frames and replacing only the glass. In either case, building managers may be understandably reluctant to replace existing windows or glass whose performance is generally adequate, though not optimum, in the case of blocking unwanted heat.

For all existing glass and in much new construction, applied window film is the least expensive and preferred solution to mitigate the impact of too much solar heat entering windows. Conventional dark and reflective applied window films successfully block a significant amount of solar heat, thereby reducing the use of HVAC systems.

Unfortunately, the same films reduce a significant percentage of visible light through the glass. Most of these films are highly reflective in daylight, giving them a mirror-like appearance when viewed externally. In artificial light and at night, internally they appear mirrored. In the case of retail establishments, visible light is reduced inside the store and shoppers outside cannot see clearly inside.

Most conventional window films transmit less than 35% of visible light, a good 35% less than the 70% necessary to be undetected by the naked eye. The result is that building interiors are correspondingly darkened, often requiring the use of increased illumination. This leads to higher electricity consumption that may increase inside temperatures requiring more air-conditioning. Increased utility costs defeat the major benefit of the film—cost savings.

THE BEST SOLUTION TO OVERHEATING—CLEAR, SPECTRALLY SELECTIVE FILM

Clear, spectrally selective applied window film offers the best ratio of visible light transmission to heat rejection. Spectrally selective refers to the ability of the film to select or let in desirable daylight, while blocking out undesirable heat.

While some manufacturers call their films spectrally selective, the definitive test is, how much visible light the film transmits. Most so-called spectrally selective films transmit no more than 54% of visible light. If a window film looks tinted and not clear, it is not optimally selective in the all-important category of visible light transmission.

The following table shows how different kinds of glass and applied films transmit light and heat.

Building management should consider the following points when evaluating spectrally selective vs. conventional window films.

☆ This entry originally appeared as "Comparing the Energy Conservation Capabilities of Spectrally Selective and Conventional Applied Window Film" in *Energy Engineering*, Vol. 102, No. 4, 2005. Reprinted with Permission from AEE/Fairmont Press.

Keywords: Window film; Applied film; Spectrally selective; Visible light transmission; Reduced HVAC operating cost.

Table 1 Window film performance

Type of glass or applied film	Percentage of daylight through glass	Percentage of solar energy through glass	Shading coefficient[a]	Luminous efficacy constant[b]	Percentage of light reflectance interior/exterior
¼″ clear glass	89	77	0.96	0.93	7/7
¼″ clear glass with tinted film	37	64	0.74	0.50	6/6
¼ clear glass with reflective film	37	44	0.51	0.73	18/28
¼″ clear glass with clear spectrally selective film	70	45	0.51	1.37	8/8

[a]The lower the shading coefficient, the lower the solar heat gain.
[b]Luminous efficacy constant, a measurement of a window glass or film's ability to simultaneously block heat and transmit light. (Visible light divided by the shading coefficient). The higher the number, the more efficiently the glass or film blocks heat and transmits light.

How do they Compare in Clarity?

The ideal film would be totally clear, yet, able to significantly block unwanted solar heat and reduce glare. Most dark and reflective films transmit less than 35% of visible light and correspondingly appear unclear. Spectrally selective film, which blocks heat equivalent to the darkest films, transmits 70% of the visible light and in so doing possesses a clear appearance (Table 1).

Data by Southwall Technologies, Inc., Palo Alto, CA, and Lawrence Berkeley National Laboratory, Berkeley, CA.

How do they Compare in Blocking Heat?

Most conventional tinted films transmit over 65% of solar energy, giving them an unacceptable shading coefficient of over 0.70 (the lower the shading coefficient, the lower the solar heat gain). With a shading coefficient as low as 0.51, reflective films block more heat, but many transmit as little as 15% of the visible light. When considering both heat rejection and light transmission, spectrally selective films out perform conventional competitors.

How do they Compare in Mitigating Heat Loss in Cold Weather?

Both conventional and spectrally selective window films are designed to block near infrared or solar heat. However, both conventional and spectrally selective window films will enhance the ability of existing glass to insulate against heat loss by as much as 15%.

How do they Compare in Applicability to Different Types of Glass?

Both conventional and spectrally selective films can be applied to single pane and insulating fixed glass, windows and doors. Always identify existing glass and follow the advice of a qualified film installer.

According to tests conducted by independent laboratories under the auspices of the Association of Industrial Metallizers, Coaters, and Laminators (AIMCAL), applied window film properly installed on insulating glass does not cause seal failure. Accordingly, most window film manufacturers offer an insulating glass warranty in the event of seal failure. For further information on the use of window film on insulating glass consult AIMCAL, Ft. Mill, SC (www.aimcal.com).

How do they Compare in Requiring Special Care?

The best-applied films require no special care. They can be cleaned just like the surface of glass using no abrasives, just soap and water.

How do they Compare in Price?

The price of dark, tinted, and reflective window film ranges from 4 to 6 dollars per installed square foot. Depending on the particulars of the installation and the geographic area, the best spectrally selective applied window film ranges in price from approximately $9 to $12 per square foot installed. Installed prices are volume dependent; on larger projects such superior performing films may be installed for less.

How do they Compare Aesthetically?

Conventional dark and reflective window film changes the appearance of existing glass and therefore the external appearance of a building. Clear, spectrally selective film does not change the appearance of existing glass, allowing its application on the entire building or on as few windows as necessary to deal with a localized overheating problem. For limited applications, spectrally selective film is competitive in price with conventional film.

How do they Compare in Payback?

Less expensive, conventional window films have a shorter payback compared to more expensive, spectrally selective films. However, it's not that simple. It is necessary to add on the cost of extra energy used for lighting due to the inability of conventional film to transmit sufficient visible light. Also, because extra lighting generates additional heat, the use of conventional window film may also increase air conditioning cost.

In reality, the payback for conventional film and spectrally selective film becomes comparable. Given rising electricity and natural gas rates, the rate of payback for spectrally selective film is always improving—averaging less than four years.

Use of Spectrally Selective Applied Window Film at Stanford University

Spectrally Selective Window Film Saves Energy at Stanford University.

While recent summers will be remembered for actual and threatened energy blackouts in California, Stanford University is doing its part to reduce energy use, thanks to the ongoing energy conservation program. Among many measures taken to improve energy efficiency at Stanford is the installation of clear, spectrally selective applied window film in 20 academic and administrative buildings on the 8,000 acre campus south of San Francisco.

At Stanford, the least expensive option to dramatically reduce unwanted solar heat and improve the performance of existing windows is applied window film. Scott Gould, Stanford's energy engineer, contends that since the 80s, window film performance has improved and that films look better and last longer. However, it was clear to decision-makers at Stanford that not all windows films are alike.

Not any film would do.

According to Alan Cummings, associate director of Facilities Operations, the primary reason for window film at Stanford is heat loading and occupant comfort. In collaboration with campus architect David Neuman, Cummings reviewed a variety of applied film products. Their primary objective was to select a window film with high light level transmission and heat load reduction. Equally important was the need for a window film that would not appear reflective. According to Neuman, the traditional architecture of the campus precludes using reflective glass, even on the newest buildings.

Conventional mirrored and tinted window films do prevent some solar heat loading, but cannot transmit high levels of light. Spectrally selective film freely transmits visible daylighting while blocking the near infrared and UV portions of the sun's spectrum. Tinted films may reduce heat gain but darken building interiors; spectrally selective film is virtually clear and so doesn't change the color of existing glass.

Over the past five years, spectrally selective applied film has been applied to selected south-, west-, and east-facing facades on 20 Stanford buildings totaling 120,000 ft^2. They include the Stanford Law School and the Green Earth Sciences Building. The film's energy payback, through lowered air-conditioning bills, is from three to four years, depending on the building, electricity rate, and weather.

The most recent film installation at Stanford took place at Encina Hall, a renovated administration building that was originally constructed as a dorm in 1891 and completely renovated in 1998. Some 6,212 ft^2 of film was applied in June, 2003.

According to a before and after energy audit conducted by V-Kool, Inc., the building has a British thermal unit meter reading (Btu/ft^2/h.) of 225 and 5.71 h a day of peak load. Daily air conditioning requirements to remove heat without the spectrally selected film amounted to 665.57 A/C tons to remove heat at a cost of AC at $66.56 per day. Daily air conditioning requirements with the film installed are 339.44 A/C tons to remove heat at a cost of AC at $33.94 per day.

Energy savings on the building with the film consist of:

1. $32.61 daily in A/C savings;
2. $978.39 in monthly A/C savings;
3. $4,891.95 in annual A/C savings.

Approximate project cost: $43,000. Return on investment: nine years or less, gave increase in the cost of electricity.

The window film installation program at Stanford University is an example of a long-term commitment to energy conservation on a campus with some buildings over 100 years old. To our knowledge, there are more buildings equipped with spectrally selective window film at Stanford University than at any other single institution in the country.

A case in point is the Los Angeles Department of Water and Power's (LADWP) rebate program for window film. It is based on a film's luminous efficacy constant, a measurement of its ability to simultaneously block heat and transmit light. While a very reflective film that blocks more heat than a spectrally selective film earns a 55 cent per square foot rebate from LADWP, a spectrally selective film that blocks less heat but lets in more light receives a higher rebate of 85 cents per square foot. Only spectrally selective films with luminous efficacy constants over 1.0 receive the higher rebate.

How do they Compare in Guarantees?

The best applied films are guaranteed not to peel, discolor, blister, bubble or demetalize for at least 10 years on a commercial installation. Look for a guarantee from the manufacturer in addition to any by the installer.

Where can I Find More Information on Conventional and Spectrally Selective Window Film?

The International Window Film Association, Martinsville, VA, (www.iwfa.com) and the AIMCAL, Ft. Mill, SC, (www.aimcal.com) provide a range of information.

REAL-LIFE INSTALLATIONS OF SPECTRALLY SELECTIVE WINDOW FILM

Both company-owned and franchise properties of the following retailers use spectrally selective window film in selected establishments: Hallmark Cards, Calico Corners, Public Storage, Albertson's, Esprit, McDonalds, Exxon, and Quik Trip convenience stores. Spectrally selective window film is saving energy at the Eldorado Country Club in Palm Springs, the Ontario, CA, convention center, and in such landmark buildings as the former headquarters of Montgomery Ward in Chicago and the headquarters of the American Institute of Architects in Washington, DC.

CONCLUSION

The universal applicability of spectrally selective applied window film makes it particularly cost effective for institutional, commercial, and residential structures in need of solutions to one or more of the following problems: solar overheating; reduction in HVAC size, operation, and cost; reduction in discomfort to building occupants and visitors due to high temperatures; increasing productivity of building occupants through maximizing natural light; utilization of floor space adjacent to windows and fixed glass; some increase in insulation against heat loss in winter; reduction of some glare while maintaining high levels of natural light; reducing internal temperature swings that can damage expensive interior landscaping; increasing a building's energy efficiency without compromising its historic and aesthetic qualities; mitigating the impact from ultraviolet radiation (UV) induced fading of furniture and window treatments; reducing exposure to cancer-causing UV, and increasing the resistance of existing glass to wind-blown debris, earthquake stress, explosions, and forced entry.

Of course, film manufacturers should and will continue research and development of applied films that will block even greater amounts of heat while transmitting high levels of visible light. Other enhancements in applied window film that may one day become available include increased durability, strength, and resistance to environmental and climatic stress, resulting in even greater film life expectancy.

BIBLIOGRAPHY

1. Watts M. Spectrally selective window film reduces heat. *College Planning and Management*, December, 2003. Case study of film use at Eastern Oregon University.
2. Barista D. Sun block, *Building Design and Construction*, December 2003. Review of latest window film technology.
3. Yates P. Seeing the light without feeling the heat. *Construction Canada*, November 2003. The benefits of energy efficient window films.
4. Abate K. Vacuum-sputtered coatings in glazing applications. *The Construction Specifier*, July 2003. Review of how the sputtering technology used in window glass and film manufacturing improves energy efficiency.
5. Yates P. Window films. *Window World*, Fall 2002. A review of the growth of window film use in Canada.
6. Egri J. Caught on film. *California Real Estate Journal*, August 5, 2002. Review of developments in window film.
7. A look at Stanford's conservation plans. *Energy & Power Management*, Coxsackie, NY, USA, October 2001. A review of Stanford's use of spectrally selective window film and other energy saving technologies.
8. Applied window film protects against heat damage to merchandise. *Energy & Power Management*, Coxsackie, NY, USA, November 2000. Reviews the use of window film by three major convenience store chains.
9. *Window Film Magazine*, Safford, VA, USA (http://www.windowfilmmag.com) is the only publication written for and about window film manufacturers, distributors, dealers and installers. It often contains material on the energy benefits and performance of applied window film.
10. *Energy Management Program Window Film Training Guide* published by the Association of Industrial Metallizers Coaters and Laminators, Window Film Committee. Ft. Mill, SC, USA. (http://www.aimcal.org) Excellent review of how window film is made and its energy conservation performance.

Windows: Shading Devices

Svetlana J. Olbina
Rinker School of Building Construction, University of Florida, Gainesville, Florida, U.S.A.

Abstract
A shading device, as an integrated component of a window and facade, protects space from direct sun, overheating, and glare; and provides increased daylight levels, desired privacy, or a view to the outside. This paper presents a classification of shading devices based on their assembly, their material, their position relative to the facade, and the control strategy used for the shading device position adjustment. The paper also introduces the decision-making framework (DMF) that can help in the selection of the most appropriate shading device for a specific building. The DMF is a tool for the analysis of the shading device performance; it is meant to be used by architects, engineers, window manufacturers, and shading device manufacturers.

INTRODUCTION

The proper design of a building with a window as a building component has the goal of providing comfort for the occupants, as well as energy efficiency for the building, by reducing the heating and cooling loads of the building. Direct sun radiation—one of the most significant cooling loads—is admitted through the windows. Therefore, direct sun radiation through the window should be prevented by the appropriate application of shading devices.[1]

The window is part of any conventional facade system, including both single-skin and double-skin facades. Windows as multifunctional systems provide thermal, visual, and acoustic comfort, and affect air quality. The thermal requirements for windows are to protect the building from heat loss in winter and heat gain in summer, and to collect energy in winter. The visual requirements for windows are to provide the occupants with a view, daylighting, privacy, and protection from glare.

The shading device is an integrated component of the window. Proper application of the shading device is especially important in curtain wall systems. Large glass areas can create a greenhouse effect, contribute to overheating, and increase cooling loads. Glass can also cause visual problems with direct and reflected glare.[2] Therefore, the application of shading devices in windows and large glass facades is necessary for controlling sunlight penetration through the glass.

Shading devices regulate heat by maximizing the reception of welcome heat in winter, and excluding excessive heat penetration in summer.[3] The shading device protects the space from direct sun and overheating in summer, reducing the cooling loads of the building by 23%–89%.[2] As a result, the appropriate use of a shading device in a window contributes to energy savings. When designing a window and a shading device for the window, the goal is to achieve a low total energy transmittance while maintaining high light transmission and good transparency.[4]

Advanced shading devices also provide daylight for the interior space. "The maximization of daylight is recognized as one of the key-goals in low-energy design."[5] A shading device as a daylighting system can redirect daylight to spaces where daylight is needed; for example, to spaces at a large distance from the window wall. Use of daylight decreases the use of artificial lighting, decreasing the following:

- Use of electricity
- Internal heat gain from lighting
- Cooling loads

This leads to energy savings for the building. The application of daylighting can decrease energy cost by 30%.

The shading device can also provide the following benefits:

- Protection from glare
- View to the outside
- Privacy
- Collection of sun energy in the double-skin facade
- Thermal insulation during winter nights

"Venetian blinds, draperies and roller shades inside single-pane, clear glass windows, reduce heat losses by

Keywords: Shading device; Shading device classification; Decision-making framework; Shading device selection; Shading device performance; Energy-efficient design; Daylighting.

25%–40% and metallic coated shades may further reduce losses by 45%–58%."[2] However, use of the shading devices in the window can obstruct the view to the outside and limit the amount of daylight that penetrates into the interior space.

The second section of this article presents a classification of shading devices based on their assembly and material, their position relative to the facade, and the control strategy used for the shading device's position adjustment. The third section explains the decision-making framework (DMF) that can help in selection of the most appropriate shading device for a specific building.

CLASSIFICATION OF SHADING DEVICES

Various shading device systems available on the market can be classified based on the shading device's assembly, its material, its position relative to the facade, and the control strategy used for the shading device adjustment.

Shading Device Assemblies and Materials

Various assemblies and materials are used for the shading device systems:

- Architectural solutions—Shading devices are an integral part of the building (e.g., overhangs, fins, brise-soleils, window setback, and light shelves).
- Window treatments—Shading devices are industrially manufactured systems (e.g., awnings, louvers, blinds, roller blinds, solar films, shades, sun screens, drapes, and shutters).[1]

The description of some of the shading assemblies follows:

- Overhangs and fins: fixed architectural shading elements, usually in the form of the balconies or projected spaces. Horizontal overhangs are effective devices for the south orientation, while the vertical fins work better for east–west orientation.[3]
- Brise-soleils: fixed architectural shading elements that consist of horizontal or vertical brise-soleil louvers. They are effective for east–west orientations.
- Awnings: consist of a frame that supports a horizontal or sloped surface on the exterior of windows. Awnings can be made of fabric, plastic, and aluminum. They can be fixed or moveable.[1]
- Louvers and blinds: consist of multiple horizontal or vertical slats. Horizontal devices are the most efficient for south orientation. Vertical devices give the best protection for east–west orientation. Slats can be either flat or curved, fixed or moveable. Louvers are exterior devices made of galvanized steel, anodized or painted aluminum, plastics, or glass (Fig. 1). Blinds are interior or between-glass devices made of painted aluminum, perforated metal, wood, glass, or plastic (Fig. 2).[1,6]

Advanced shading devices not only meet thermal performance requirements but also improve daylight levels in the space. The examples of advanced devices are as follows:

- Light shelves: flat or curved elements that reflect light either outside (exterior light shelf) or inside (interior light shelf). They divide the window into two areas; the upper area provides daylight, while the lower area provides shading.[7]

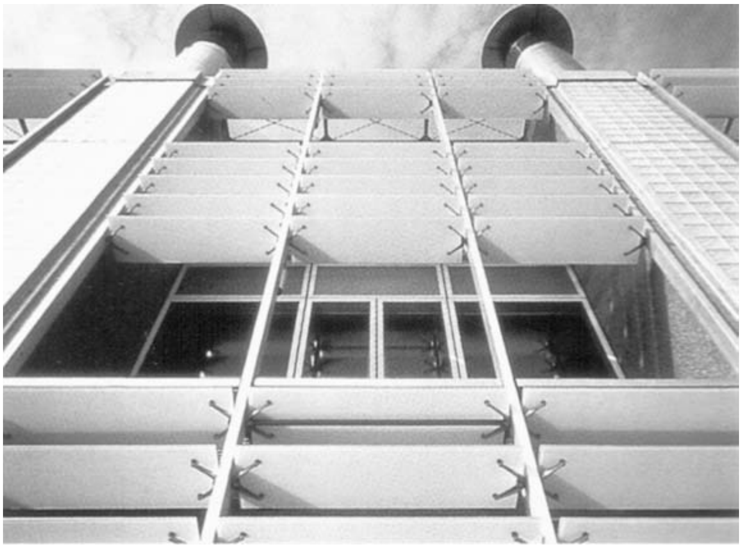


Fig. 1 Louvers: an example of the automatically controlled exterior shading device, made of glass. Source: From Architectural Press (see Ref. 5).

Win–Wire

Fig. 2 Venetian Blinds: typical example of the manually controlled interior shading devices. The slats are made of perforated aluminum.

- Mini light shelves: for example, Okasolar units have a concave and convex shape and are made of a highly reflective light-gauge steel.[8,9] Louvers are fixed at a predetermined angle and spacing to respond to different seasonal conditions. Louvers are installed between the two panes of glass (Fig. 3).
- Prismatic and refraction elements: can be made of acrylic (Fig. 4). They can be installed in the upper part of the window to protect the space from glare and veiling reflections, while the lower part of the window provides the view.[8]

Position of Shading Devices Relative to the Facade

Based on their position relative to the facade, shading devices can be classified into three major groups:

- Exterior devices
- Interior devices
- Between-glass devices

An exterior shading device is installed in front of the facade (Figs. 1 and 4). Examples include overhangs, fins, awnings, louvers, brise-soleils, fabric blinds or screens, and roller blinds.[9] Exterior devices provide better solar

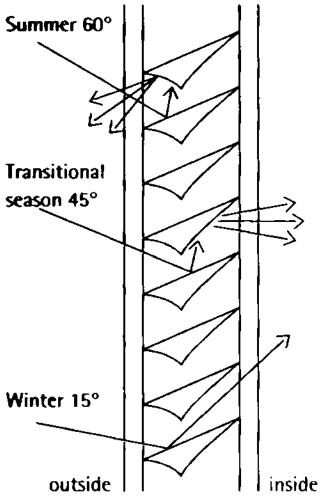


Fig. 3 Mini light shelves: the Okasolar unit is an example of the fixed, between-glass shading device, made of a highly reflective light-gauge steel.
Source: From Birkhauser Publishers (see Ref. 8).

protection in summer than interior devices, because exterior devices block sun radiation before it enters the glass panel and interior space. Shading effectiveness increases 35% by using an exterior shade instead of an interior one. As a result, building cooling loads are reduced. The exterior shading device's maintenance is more complicated and expensive than maintenance of the interior shading device. Exterior devices are more expensive because of the structural and durability requirements.

Interior shading devices are installed in the building's interior space. Examples include Venetian blinds (Fig. 2), traditional roller shades, drapes, and blackout screens.[7] Interior shading captures sun energy that can be used in

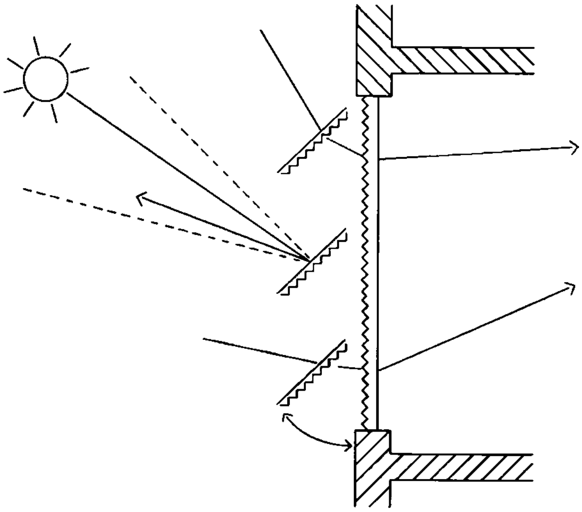


Fig. 4 Prismatic shading elements: an example of a moveable exterior shading device made of acrylic.
Source: From Birkhauser Publishers (see Ref. 8).

Win–Wire

winter for space heating. Maintenance of interior devices is easier and less expensive than for exterior devices. Efficient solar protection in summer is difficult to achieve with an interior device, since sunlight enters the interior space and overheats the space between the blinds and the interior glass layer.

A between-glass shading device can be installed in an air cavity in one of two ways:

- Between two panes of glass in a double insulating glass unit (DGU) (Fig. 3).
- Between two facade layers in a double-skin facade.

The between-glass devices are less exposed to dust and dirt, so there is less need for cleaning. If the blinds are moveable and fully automatically controlled, maintenance can be complicated and more expensive, and a more complex window structure can be required. In summer, the between-glass shading device in the DGU usually reflects all sun energy in order to protect the interior from overheating. Any energy absorbed by the shading device contributes to the heating of the glass panes and of the air in the cavity between glass panes. This creates a problem in the DGU. However, in a double-skin facade, this warm air can be exhausted at the top of the facade, so that the facade and interior space can be protected from overheating in the summer.

Control Strategy for the Shading Device Position Adjustment

Shading devices can be either fixed (Fig. 3) or moveable. "Fixed systems are usually designed for solar shading, and operable systems can be used to control thermal gains, protect against glare, and redirect daylight."[6] The use of fixed blind systems requires higher energy consumption than moveable blind systems.[10] Moveable systems follow the dynamic exterior thermal and luminous conditions.[1] Position of the moveable shading device can be adjusted manually (Fig. 2) or automatically (Figs. 1 and 4), depending on the sun position, the sun radiation intensity, and the requirements for interior temperature and light levels. The moveable shading device can have three basic positions: open, partially open/partially closed, and completely closed.

Manually operated systems are generally low energy-efficient, because occupants may or may not operate them "optimally."[6] Occupants very often close the blinds completely to protect the space from overheating and glare, but at the same time the amount of daylight in the space is reduced; therefore, both the use of electric lighting and, thus, the cooling loads are increased. "If the blinds are open when a large amount of solar radiation enters, excessive energy is consumed for air-conditioning… When the blinds are closed on days without solar radiation, the advantage of the view from the window is lost."[11] The occupants will adjust the blinds to protect the space from direct sunlight, but will rarely adjust the blinds again when the direct sunlight is gone, and daylighting can be admitted.[10]

The automated shading device systems optimize energy use and control interior conditions without relying on occupants.[7] Automated systems can achieve savings in both cooling loads and lighting energy.[12] Automated blinds have better thermal and daylighting performance than both fixed blinds and manually controlled blinds. Automated systems "close automatically when the interior becomes too glary or too hot, and re-open later to admit useful light."[10] The use of automated Venetian blinds decreases the energy cost by 30% during the winter and by 50% during the summer. However, automatic systems can produce discomfort in occupants who dislike the feeling of not having personal control over the system.[6] Automated devices are often high-maintenance, and therefore expensive, solutions.[8]

SHADING DEVICE SELECTION

Several criteria should be considered when selecting the most appropriate shading device among the devices available on the market. To make the proper choice of the shading device, the required or desired performance for the shading device and the variables that affect that performance need to be determined. Fig. 5 shows the structure of the DMF for the shading device selection. The user of this DMF can be an architect, engineer, windows manufacturer, or shading device manufacturer. The DMF is an analysis tool that can help its user to select the most appropriate shading device for a building.

The structure of the DMF includes the following:

- Independent, dependent, and shading device variables that influence shading device performance.
- Performance parameters (thermal, visual, acoustic, aesthetic, cost, and control) that are used as criteria for the shading devices' evaluation and selection.
- Relationships and interactions among the variables and performance parameters.

Independent Variables

Independent variables such as climate, location, site, and building type are given to the user of the DMF.

The United States has four major climate zones: hot dry, hot humid, cold dry, and cold humid. For each of these climates, characteristics need to be determined. Climate directly affects the type of heat transfer, Heating, Ventilation, and Air Conditioning (HVAC) conditions, the facade type, shading device variables, the shading device's thermal and visual performance, the operational cost of

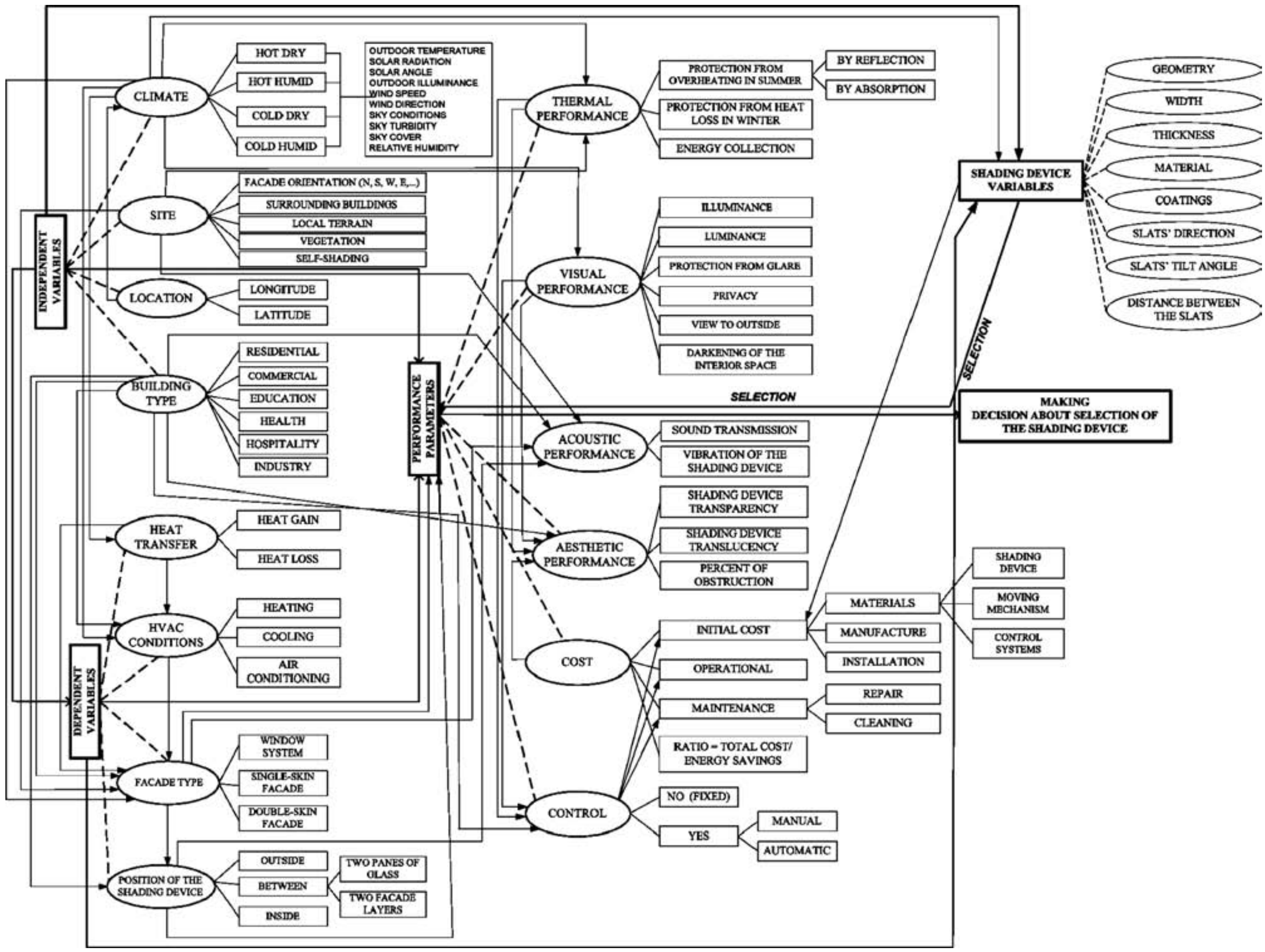


Fig. 5 The decision-making framework (DMF) for the shading device selection: the figure shows the structure of the DMF including variables (independent, dependent, and shading device variables), performance parameters (thermal, visual, acoustic, aesthetic, cost, and control), and the relationships and interactions among the variables and performance parameters.

the shading device, and the control strategy for the devices' adjustment.

The building location is defined by latitude and longitude. Based on the building location, the climate of the region and the microclimate of the particular building's site can be determined. Location indirectly affects heat transfer, HVAC conditions, the facade type, the shading device's thermal and visual performance, and the control strategy.

The building site is given to the designer, who has two choices: to select the position of the building on the site, or to accept the predefined position of the building, as in the case of the dense, urban setting. The site has a strong relationship with the climate and location. The site has direct influence on the facade type, the shading device variables, the shading device's thermal and visual performance, the operational cost, and the control strategy for the blinds' adjustment.

The building type defines a function of the building; for example, residential, commercial, education, or health. The building type directly influences heat transfer, HVAC conditions, facade type, the position of the shading device, the shading device variables, and the performance of the shading device.

Dependent Variables

The user of the DMF defines dependent variables, such as heat transfer, HVAC conditions, the facade type, and the position of the shading device relative to the facade.

The user of the DMF defines dominating heat transfer conditions in the building, such as heat gain or heat loss. Heat transfer depends on the climate, location, and building type. Heat transfer affects the HVAC conditions, the facade type, the shading device variables, and the values of the performance parameters.

By selecting the HVAC conditions, the user of the DMF decides whether or not there is a need for heating, cooling, or air conditioning—or a combination of any of these systems. The selection of the HVAC conditions is made based on the heat transfer in the building, the climate, and the building type. The HVAC conditions have an impact on the shading device variables, the shading device's thermal performance, and the control strategy for the devices' adjustment.

The selection of the facade type is affected by the climate, site, and building type. The facade type has an effect on the shading devices' variables and on the shading device's thermal, visual, aesthetic, and cost performance. For example, if the double-skin facade is chosen, the shading device can be installed between the two facade layers and function as a solar collector, thus improving the facade's thermal performance.

The position of the shading device is dependent on the climate, site, building type, heat transfer, HVAC conditions, and facade type. The position of the shading device strongly influences the shading device variables. For example, an exterior device should be made of different material than an interior device because exterior devices must be weather resistant. The position of the shading device affects the performance of the shading device, especially the maintenance cost (e.g., it is more expensive to clean an exterior device than an interior one).

Shading Device Variables

This DMF includes the following shading device variables: the shading device's geometry, width, and thickness; applied materials and coatings; and in the case of Venetian blinds, distance between the blinds, the blinds' direction, and the blinds' tilt angle. Shading device variables are independent in the process of the shading device's selection because the shading device variables are already defined by the manufacturer and given to the designer of the building. These predefined variables are used to analyze the performance of the shading device. Shading device variables, together with the independent and dependent variables, directly affect the performance of the shading device.

Performance Parameters for the Shading Devices

Performance parameters considered in this DMF are the thermal, visual, acoustic and aesthetic performance; the cost of the shading devices; and the control strategy for the shading devices' position adjustments. The performance parameters' values depend on the independent, dependent, and shading device variables. There are also interactions among the performance parameters in this DMF.

The thermal performance of the shading device includes protection from overheating in summer, protection from heat loss in winter, and collection of sun energy. Thermal performance strongly depends on the climate, site, building type, heat transfer, facade type, position of the shading device, and shading device variables. In a hot climate, the office building's shading device will be required to provide protection from overheating. The shading device's geometry, materials, and position will be chosen to achieve the required protection from overheating. The level of protection from heat loss during winter nights or in cold climates and the collection of sun energy are also measures of the shading device's thermal performance. The device can absorb solar energy instead of reflecting it, and also collect this energy for application in the building's mechanical systems. There is a strong relationship between the shading device's thermal performance, its visual performance, and its control strategy. The shading device can be designed to provide maximum overheating protection but also to allow the sufficient daylight level in the space. To achieve this goal in the case of Venetian blinds, control systems should provide the blinds' optimum tilt angle.

The visual performance of the shading device includes providing the following desired effects:

- Illuminance
- Luminance
- Protection from glare
- Privacy
- Darkening of the space
- Visual contact to the outside space

Climate and site affect illuminance, luminance, and protection from glare. The building type strongly affects the requirements for providing privacy, darkening of the space, and direct visual contact to the outside space. For example, providing privacy and darkening of the space is often desirable in residential buildings, but not necessarily in office buildings. The visual performance of shading devices depends on the facade type, the position of the device, and the devices' geometry, dimensions, material, direction, and tilt angle. Also, there is the interaction among the thermal, visual, and aesthetic performance, and the control strategy. When selecting the shading device, the user of the DMF needs to understand that good visual performance can be achieved only with a thoughtful control strategy of the shading device's adjustment. The user of the DMF also must consider the impact of such a shading device on the appearance of the facade, i.e., on the aesthetic performance.

Acoustic performance parameters of the shading device include sound transmission and vibration of the blinds. The acoustic performance is significantly affected by the following:

- The building location and site: a higher level of noise occurs in urban areas; therefore, the shading device has to be designed to reduce this noise.
- The building type: different levels of acoustic comfort are required in different types of buildings, and a

specific shading device can help in meeting the acoustics requirements.
- Facade type and position of the shading device: the exterior shading device can vibrate because of wind, resulting in increased noise level. For that reason, structure of the exterior device should be designed properly to avoid the problem of vibration.
- The thermal performance: the shading devices that protect space from heat loss during cold nights can also protect the space from noise, because devices can be made of materials with good thermal and acoustical properties.
- Control strategy: completely closed blinds provide the lowest sound transmission, but at the same time they block the daylight. The control strategy needs to balance the sound transmission, penetration of the daylight, and protection from heat loss or gain.
- The shading device variables: the shading device's material, shape, dimensions, and position need to be selected to achieve the best possible protection from noise.

The shading device has a significant impact on both the exterior and interior appearance of the facade. To achieve an aesthetically pleasing look for the shading device and facade, the device's transparency or translucency and the percent of the window area obstructed by the device need to be considered. The shading device's aesthetic performance is affected by the climate, site, and building type. The appearance of the surrounding buildings also affects the appearance of the analyzed building, and, consequently, the appearance of the shading device. Different aesthetic requirements are imposed for the shading devices installed on different building types. The shading devices can have a different look if they are installed on an office building than they would on a hospital or industrial building. The aesthetic performance of the shading device is influenced by the facade type; the shading device's position; its thermal, visual, and acoustic performance; the cost of the shading device, and the applied control systems. The shading device's transparency or translucency and the percent of the window area obstructed by the shading device depend on the requirements for the protection from overheating, the desired light level in the interior space, and the choice of the shading device's variables.

Total cost of the shading device consists of initial, operational and maintenance cost. Cost analysis should also include calculation of the ratio between total cost and energy savings achieved due to use of the shading devices. The initial cost of the shading device includes the cost of material, cost of manufacture of the system, and cost of installation of the shading device on the facade. The initial cost includes also the cost of the moving mechanisms and control systems if the device is moveable. The operational cost includes the operating cost of the shading device itself and the cost of the heating, cooling, and lighting of the space, which is a result of the application of the shading device on the building. The cost of maintenance includes cleaning and repair costs. The cost of the device is influenced by the climate; location or site; building type; thermal, visual, acoustic, and aesthetic performance parameters; and control systems.

The control strategy for the shading devices' adjustment in this DMF considers two options: shading devices without need for control (fixed) and shading devices controlled manually or automatically. The control strategy for the shading devices' adjustment depends strongly on the following:

- Climate, particularly sun radiation, sun angle, and sky conditions
- The building type, because different building types require different values of the performance parameters for the shading device; and, therefore, different positions and tilt angles of the device
- The facade type
- Heat transfer
- HVAC conditions
- The shading device variables, particularly geometry of the device, dimensions, and materials
- The required thermal, visual, and acoustic performance, and the cost of the shading device

Making the Decision by Using the DMF

The process of making a decision about selection of the shading device by using this DMF includes the following steps:

- Identifying input for the DMF—The user of the DMF prepares the input for
  - The independent, dependent, and shading device variables
  - The required values of the performance parameters that can be taken from active standards, codes, and recommendations
- Testing the shading devices—Separate testing is performed for each type of shading device to analyze thermal, visual, acoustic, aesthetic, and cost performance; and the effect of the applied control strategies. Depending on the nature of the performance parameter, the actual values of the performance parameters can be obtained by experimental testing, computer simulations, and mathematical calculations.
- Obtaining output results of testing—The actual values of thermal, visual, acoustic, and aesthetic performance parameters; the cost of the shading device; and the effect of the control strategy are collected for each type of shading device. The results are organized in an

understandable and useful format and prepared for analysis.

- Making the decision about the selection of the most appropriate shading device for the particular building—Output results of testing are compared to the required values of performance parameters for the shading devices. If the shading device's actual performance meets the requirements of the standards, then the particular device can be considered for further analysis and application on the building. Then the alternative shading devices are compared to each other. The actual values of the performance parameters for each shading device are compared, and the device with the best overall performance is selected for use on the specific building.

CONCLUSION

Shading devices are integrated elements of windows and building facades. They increase the energy efficiency of buildings and improve comfort for the building occupants. Shading devices provide the following benefits:

- Protection from direct sun
- Protection from overheating
- Protection from glare
- Increased daylight levels
- Privacy
- View to the outside space

Different types of shading devices that exist in the market can be classified based on their materials, their assemblies, their position relative to the facade, and the control strategy for the shading device position adjustment. The DMF presented in this paper helps the user, whether an architect, engineer, window manufacturer, or shading device manufacturer, to select the most appropriate shading system among several available systems. The DMF offers the user a tool for an analysis of the shading device's performance.

REFERENCES

1. Schuman, R.; Rubinstein, F.; Papamichael, K.; Beltran, L.; Lee, E.S.; Selkowitz, S. *Technology Review: Shading Systems*; Lawrence Berkley National Lab: Berkley, CA, 1992.
2. Dubois, M. *Solar shading for low energy use and daylight quality in offices. Simulations, measurements and design tools*, Report No TABK-01/1023; Department of Construction and Architecture, Lund University: Lund, Sweden, 2001.
3. Olgay, A.; Olgay, V. *Solar Control and Shading Devices*; Princeton University Press: Princeton, NJ, 1957.
4. Koster, H. *Dynamic Daylighting Architecture: Basics, Systems, Projects*; Birkhauser Publishers: Basel, SW, 2004.
5. Wigginton, M.; Harris, J. *Intelligent Skins*; Architectural Press: Oxford, UK, 2002.
6. Ruck, N.; Aschehoug, Ø.; Aydinli, S.; et al. *Daylight in Buildings: A Source Book on Daylighting Systems and Components*, A Report of IEA SHC Task 21/ECBCS Annex 29; Lawrence Berkeley National Laboratory: Berkley, CA, 2000.
7. Cardomy, J.; Selkowitz, S.; Lee, E.S.; Arasteh, D.; Willmert, T. *Window Systems for High-Performance Building*; W.W. Norton & Company: New York, 2004.
8. Daniels, K. *Low-Tech Light-Tech High-Tech*; Birkhauser Publishers: Basel, SW, 1998.
9. Campagno, A. *Intelligent Glass Facades: Material, Practice, Design*; Birkhauser Publishers: Basel, SW, 2002.
10. Galasiu, A.D.; Atif, M.R.; MacDonald, R.A. Impact of window blinds on daylight-linked dimming and automatic on/off lighting controls. Solar Energy **2004**, *76*, 523–544.
11. Inoue, T.; Kawase, T.; Ibamoto, T.; Takakusa, S.; Matsuo, Y. The development of an optimal control system for window shading devices based on investigations in office buildings. ASHRAE Trans. **1988**, *104*, 1034–1049.
12. Lee, E.S.; DiBartolomeo, D.L.; Selkowitz, S.E. Thermal and daylighting performance of an automated venetian blind and lighting system in a full-scale private office. Energy and Buildings **1998**, *29*, 47–63.

Wireless Applications: Energy Information and Control

Jim Lewis
Obvius LLC, Hillsboro, Oregon, U.S.A.

Abstract
Most owners and managers of commercial and industrial facilities recognize the value of using interval energy information to allocate costs to building users and fine-tune building operations, but find that installation costs are very high in existing buildings. Some of the most promising technological developments in the submetering world are centered on the use of wireless devices to gather data from the meters, communicate it to a central data acquisition location, and forward the interval data to a remote server using wireless Web networks. This entry examines some of the key elements of wireless energy information and control networks and provides an overview of important underlying wireless infrastructures. Particular emphasis is placed on using "mesh" radio networks to communicate raw meter data within a building or campus and using GSM/general packet radio service (GPRS) cellular networks for long-haul communications.

INTRODUCTION

Recent volatility in energy prices has made submetering (i.e., any meter located on the customer side of the primary utility meter) a valuable tool for building operations and management. One of the biggest obstacles to getting this kind of information has been the high cost of running wiring from the meter hardware (gas, electricity, water, etc.) to a data acquisition device in existing buildings. Many radio manufacturers (such as Maxstream and AeroComm) have introduced low-cost radio modules that can be adapted to provide the backbone of a wireless metering system. Using these off-the-shelf radio modules in conjunction with specially designed hardware and software for RS485 communications provides a seamless wireless communications network for both new and existing meters. In addition to this internal communications network, the development of GSM/general packet radio service (GPRS) cellular networks for wireless networking (think personal digital assistants (PDAs) with email capability) provides a convenient means of gathering interval data from locations around the world without the need for connecting to an existing local area network (LAN) or phone line.

A SUBMETERING PRIMER

For most end users, when we use the term "submetering," the immediate image is of the hardware installed on the electrical system to measure kW and kWh. While these devices are certainly an indispensable part of a submetering system, they are only one of the components. A successful submetering system takes the raw data from one or more meters and converts it into useful and timely information to be used for the following:

- Cost allocation to tenants
- Operational analysis and improvement
- Measurement and verification of energy savings
- Benchmarking and accountability

Converting the raw data at the meter level to actionable information requires several steps. Within the metering industry, this process can best be illustrated by looking at five key components:

1. *Meters and sensors*—This is the hardware level where the devices are actually installed in the electrical (or gas or water) system to capture energy consumption for one or more different areas of the building. In the electrical system, these meters might be the traditional round glass meters or they might be meters designed especially for submetering.
2. *Internal communications*—This is the mechanism used to communicate data from the meters and sensors to the data acquisition server (DAS). In traditional applications, this is usually a simple twisted pair of wires connected on an RS485 serial daisy chain.
3. *Data acquisition*—The meters and sensors from level 1 produce industry standard outputs (e.g., pulse or Modbus) that correspond to the real-time outputs being monitored. For example, each pulse from an electrical or flow meter has an assigned value (e.g., 1 pulse = 10 kWh). In order for this raw data to be utilized, it must be captured in a timely manner, time-stamped, and made available for presentation.

Keywords: Submetering; Wireless; RS 485; Modbus; Energy information.

Win–Wire

The DAS takes this raw data, time-stamps it, and then sends it to a local or remote server for storage and report generation.

4. *External communications*—Once the data is collected by the DAS, it is sent to a local or remote server for reporting. In traditional metering applications, this communication may use a LAN connection or a modem.
5. *Storage and reports*—Once the data from the meters is gathered at the building level, it still requires further processing to produce user-friendly reports and information. This processing occurs at the user interface level, where information from one or more buildings is collected into a traditional database for reporting and storage.

A detailed discussion of each of these components is beyond the scope of this entry, but most successful submetering systems have the following characteristics:

- "Open" meter protocols that provide the capability for any DAS to connect to any meter or sensor, regardless of what is being measured (e.g., electricity, gas, water, flow, british thermal units (BTUs)). Most meter manufacturers use Modbus remote thermal unit (RTU) as the primary protocol for communications with submeters, but the DAS should also accept pulse or analog values that meet industry standards.
- Nonvolatile storage of data—the DAS should be able to store at least 30 days of data from multiple meters without the risk of data loss in the event of power failure.
- Options for user-selected data intervals—the user should be able to select data intervals to match the utility interval (e.g., 15 min).
- "Open" protocol support at the DAS level that allows the data collected to be sent to any Web server (i.e., nonproprietary format). In most cases, this can be accomplished using standard Internet protocols such as HTTP, FTP, and XML. This allows the end user to select virtually any software program (including custom programs), regardless of the DAS or meter brand.

Fig. 1 (below) shows a graphical representation of a traditional wired submetering system, starting with the meters and sensors at the bottom and moving up to the storage and reports level at the top of the figure.

WHY DO WE NEED WIRELESS SUBMETERING?

As outlined in the previous section, there are readily available wired solutions that provide reliable submetering data using commonly available tools such as serial ports and LANs—so why do we need wireless? The simplest answer is that in many applications (particularly retrofits), the cost of running wire between meters and the DAS is prohibitively expensive, and using wireless components can dramatically reduce both the cost and the installation time. In addition, in many submetering projects, the cost and time involved in securing a network connection or phone line for communicating from the DAS to the Internet becomes a major headache.

Using wireless solutions in submetering not only minimizes the costs of many projects, but it also greatly reduces the disruption of day-to-day operations caused by running wire several hundred feet or more through an existing facility. Wireless also provides a very attractive alternative to trenching between buildings in a campus environment and eliminates the need for coordinating with the IT department for use of an existing LAN connection.

WHICH "WIRELESS" ARE WE TALKING ABOUT?

Ask anyone what the term "wireless" means, and you'll probably find yourself on the receiving end of one of those looks that implies you must be one of the dumbest people on Earth. Wireless, as any fool knows, means no wires, and therein lies the problem with understanding wireless submetering. A short list of wireless terms includes the following:

- Satellite
- Cell phone
- Wi-fi
- Pager
- Proprietary radio
- Bluetooth
- Zigbee

In essence, wireless (or for that matter, wired) communications form the bridge between each of the levels of metering discussed above. We will be examining two different wireless applications in this entry, the first being communications between meters and the DAS and the second the communications of interval data from the DAS to a remote server for storage and reporting.

PART I-METER-LEVEL WIRELESS COMMUNICATIONS

One of the most expensive aspects of adding submetering to an existing facility or facilities is the cost of running wires from multiple locations to the central data acquisition point. This cost can be significant, particularly when wiring must be run between multiple buildings (i.e., campuses or bases). The first level of wireless communications we will consider is the communication of data from meters and sensors wirelessly to the DAS. This

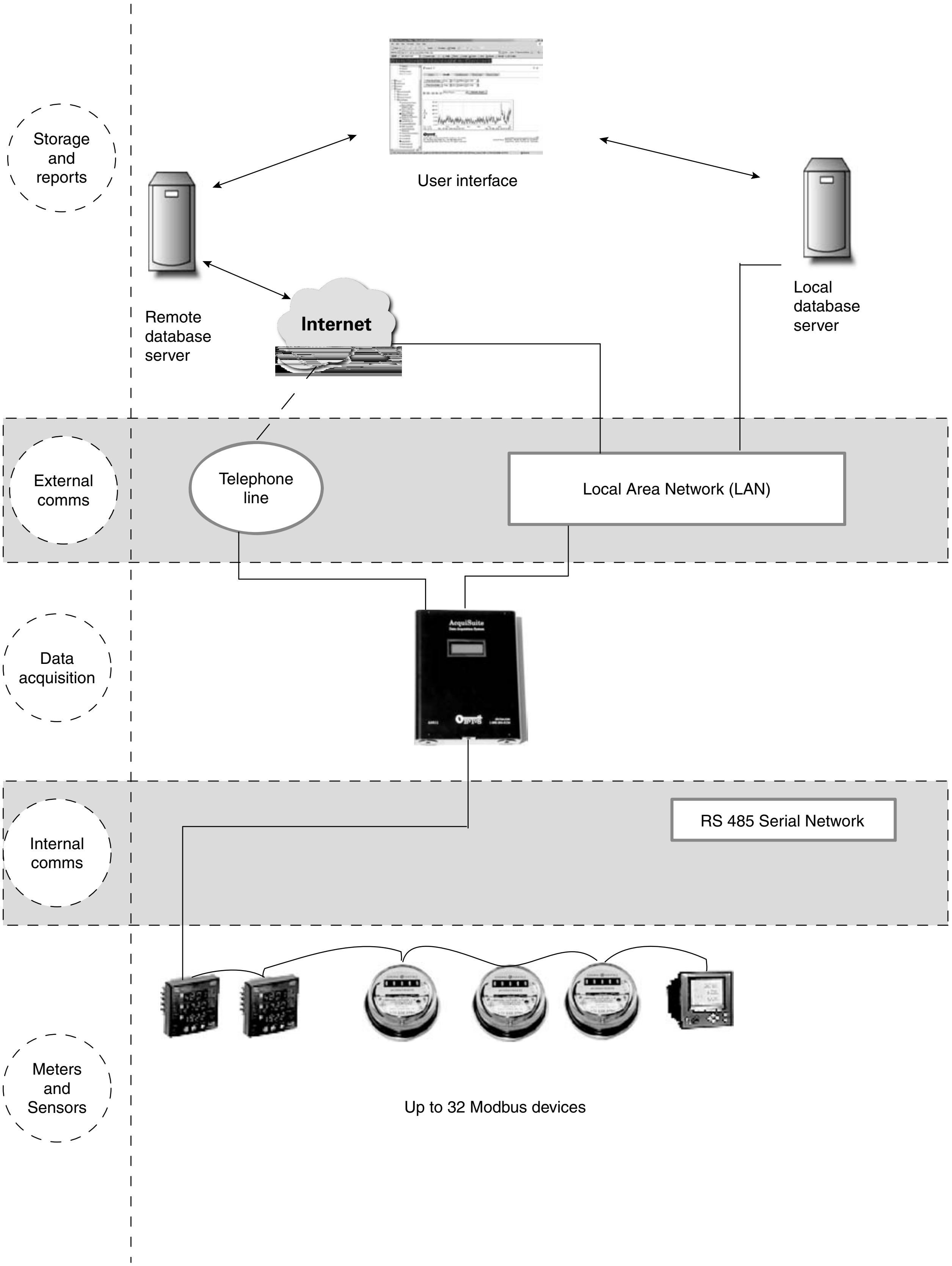


Fig. 1 Overview of metering system components.

level represents the area in which technology is advancing the most rapidly, as electronics developed for a wide variety of wireless communications are being utilized to reduce the cost of submetering.

A BRIEF PRIMER ON WIRELESS TERMS

A detailed technical discussion of wireless technologies is beyond the scope of this entry, so the purpose here is just to provide a brief overview that will serve as a jumping-off point for later discussion. At the most basic level, all wireless communications are based around radio waves, not unlike the over-the-air broadcast of radio and TV signals. (For the younger readers, yes; it is true that in the olden days, TV was limited to a half-dozen channels captured out of the air with a pair of rabbit ears.) There are significant differences between the radios used, but at the heart are some shared attributes that help to determine the suitability for use in wireless metering. Among the most common and important terms associated with radios are the following:

1. *Frequency*—Frequency refers to the time between peaks of the radio signal generated by the radio. In order for two radios to communicate with each other, they must be tuned to the same frequency, whether in a broadcast (one-way) or two-way communication environment.
2. *Power*—Most radios are rated for specific power levels that serve to determine how far a signal can be detected.
3. *Interference*—As anyone who has ever driven under a power line or tried to use a cell phone knows, there are a number of things that can cause interference with a radio signal, and the same is true for wireless metering technologies. Concrete walls, steel panels and other radio sources all provide challenges to the successful communication of metering data.
4. *Throughput*—A combination of the above factors, plus a few others, determines the throughput of the radio system. Basically, this just means how much data can be successfully transmitted in a given period of time, including any repeat requests or other delays.
5. *Repeaters*—Despite the best design of radios, it is very likely that there may be "dead spots" or areas where radio transmission is not successful. This may be due to interference, lack of power, or some other factor; but regardless of the cause, it may be necessary to install repeaters, which are basically designed to relay a lost or weak signal from one point to another.
6. *Mesh networks*—The term "mesh network" refers to networks designed to be self-healing radio networks that automatically configure themselves to optimally send data to one or more other radios without the need for setup and configuration by the user.
7. *Magic*—One of the most interesting characteristics of most radio networks is that they represent a blend of science and art. Most cell phone users have experienced the joy of having a cell phone that always works in a particular location, but suddenly doesn't, and the same magic applies to radios used for metering applications. While there are many tools available to assist in predicting the odds of success in any given location, the reality of wireless communications is that even the best of tools may not prove 100% reliable.

WIRELESS SERIAL COMMUNICATIONS NETWORKS

In order to better understand some of the options available for wireless communications between meters and the DAS, let us take a closer look at the wired version of this same communication. Fig. 2 shows a section of the previous figure that focuses on the internal serial communications networks for submeters.

As Fig. 2 shows, in a wired sensor-level communications network, a twisted pair of wires is run from a serial port (RS485) on the DAS and daisy chained to each of the Modbus meters or devices. The DAS uses plug-and-play technology to detect the type of meter and load the appropriate drivers to interrogate the meter(s) to obtain the desired data. On preselected intervals (typically 15 min or 1 h), the DAS will gather data from the meters that may be limited to consumption (kWh) or may include other parameters such as power factor or total harmonic distortion (THD), if the meter supports those additional functions. The readings are then stored by the DAS until they are "pushed" or "pulled" to a remote server for storage in a standard database (more on that later).

This system is quite reliable and provides an excellent method of communication if the meters and the DAS are all in close proximity. This type of serial network is typically capable of reliable communications up to 4000 ft. On the other hand, if there are multiple locations throughout the facility or campus that make wiring difficult or expensive, wireless nodes near the meters can provide a very cost-effective alternative to wiring.

HOW DOES WIRELESS SERIAL COMMUNICATION WORK?

There are at least a couple of different ways of wirelessly transferring data from meters. One method involves using

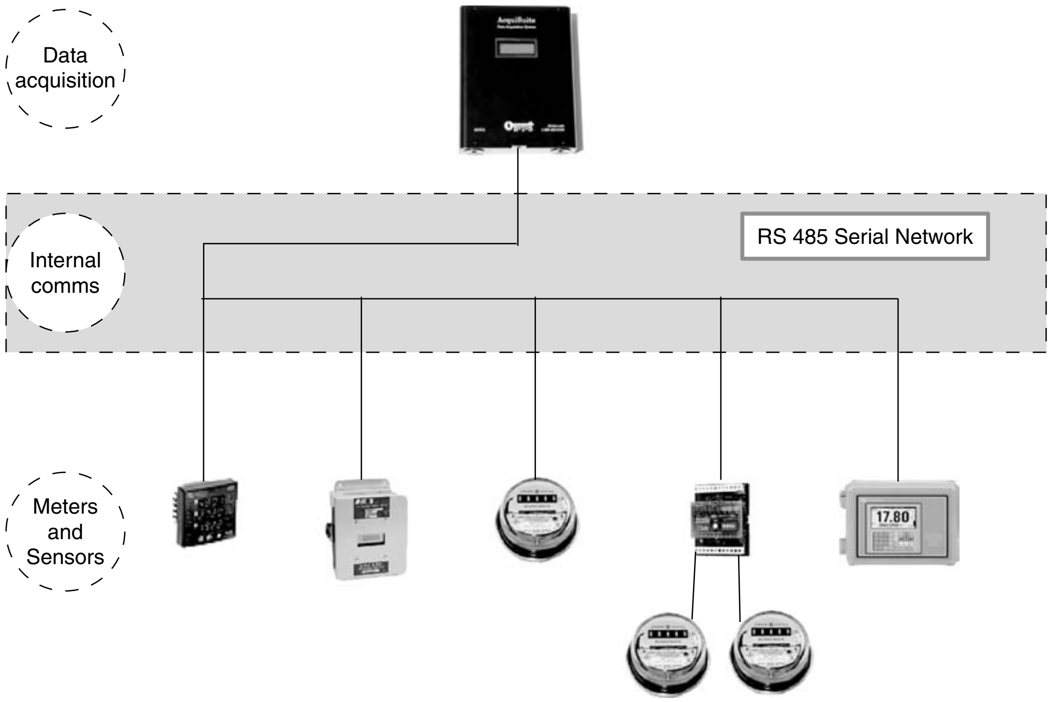


Fig. 2 Internal metering components view.

low-powered battery devices to capture and transmit pulses from meters to the DAS, which represents a very low-cost means of communications, but is not a replacement for the ideal serial network shown above. The wireless nodes function in one-way communication only and cannot provide the reliability or the additional data available from reading and sending Modbus serial data to the network.

The second (and preferred) method is to simply use wireless nodes on either end of the transmission to seamlessly replace the twisted pair of wires in Fig. 2. The newest technologies on the market accomplish this in a mode that is completely transparent to the user (i.e., the data transmitted wirelessly is indistinguishable from the data gathered on a wired network).

Fig. 3 shows a typical system that substitutes wireless Modbus communications for wired networks. In this case, we will assume that the two meters on the left of the graphic are located in one electrical room; the three on the right are located in a second electrical room; and the single meter in the center is the primary meter for the facility, which is located in close proximity to the DAS, and thus can be hard-wired.

In this example, all of the meters can communicate all of their data using the standard Modbus protocol, regardless of whether they are connected wirelessly or via wires. In the case of the remote meters, we have simply wired the RS485 outputs from the meters into a wireless node that puts the Modbus data into a wireless form and sends the data via radio to another node connected directly to the DAS, where the data is once again converted to RS485.

SO HOW DOES THIS MESH NETWORK FUNCTION?

The two keys to making the wireless communications network shown in Fig. 3 actually work are (1) that the individual radio transceivers (or nodes) are specifically designed for transmitting Modbus data and (2) that the radios contained within the transceivers can function as part of a "mesh" network. A brief overview of the functioning of a mesh network will provide a better understanding of the benefits of using wireless mesh networks to replace wired serial communications.

All mesh networks (regardless of the application) share certain common characteristics that make them valuable for replacing or augmenting wired networks. Among the most important are:

- *Self-configuring nodes*—The whole concept of mesh networks is that the radio nodes will be aware of other nodes and will automatically configure the network to optimize throughput of data, without the need of programming or configuration on the site.
- *Multiple routing paths*—Because each node is "aware" of all the other nodes it can communicate with, there are multiple options available for routing the data. As new nodes are added to the network, the network should recalculate the optimal routing paths to take advantage of more efficient routing options.
- *Spread spectrum radios*—This feature allows the radios to work on a variety of predetermined frequencies, minimizing the likelihood of interference

Win–Wire

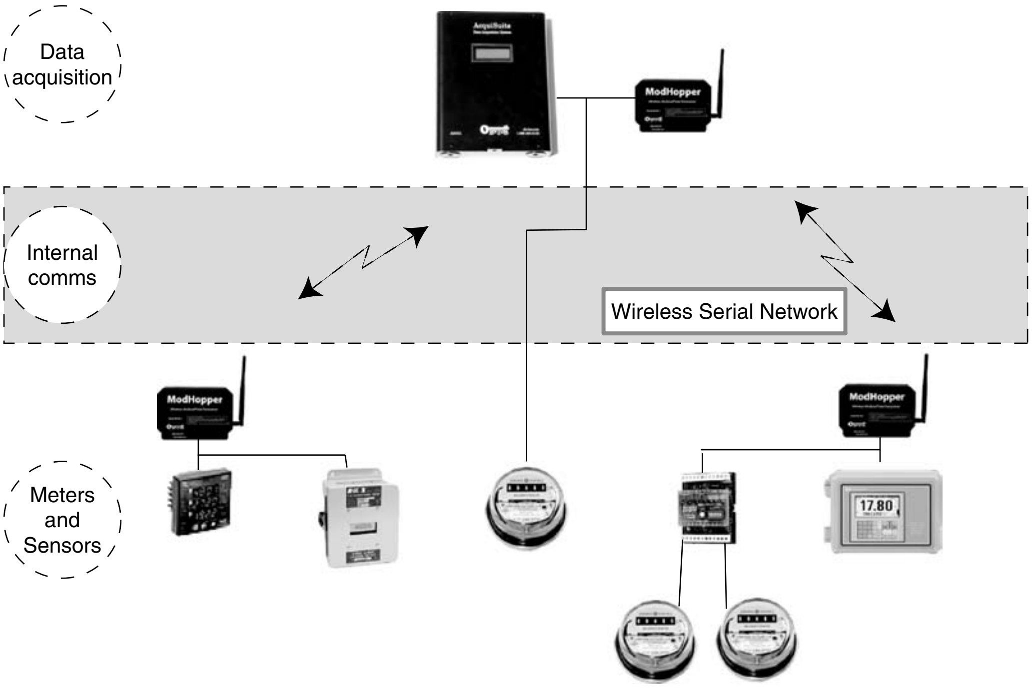


Fig. 3 Metering components with wireless communications.

from other radio sources. Essentially, the radios are designed to "hop" to different frequencies if interference is encountered, until the two radios find a frequency that minimizes interference.

- *ISM band radios*—Transceivers that use the instrumentation, scientific, and medical (ISM) frequencies set aside by the FCC do not need to have a local license for each site. These radios are designed to share the frequency bands with other systems and devices with a minimum of interference.

Fig. 4 shows a diagram of a typical wireless mesh network. Each of the transceivers can accept up to 32 Modbus devices and 2 pulses.

This figure shows a mesh network with eight transceivers and nine meters, all connected to a DAS for data collection. Two of the transceivers are functioning as repeaters, because they have no meters connected and serve only to route traffic to other nodes. The dashed lines between the transceivers show the other nodes that each can communicate with, and most of the nodes can see multiple other nodes, providing multiple routing paths for requests for data from the DAS. The network will automatically figure the optimal routing to reach each of the meters and will adjust these routing paths based on the success rate of transmissions. If other Modbus devices are added to any of the nodes, their presence is detected by the network and optimal routing paths are determined. If other transceivers are added, they will similarly be detected by the network and added to the routing paths.

SUMMARY

The technology available today makes wireless submetering a viable alternative to hardwired systems. The installation of these products, whether as the total solution or in conjunction with wired Modbus meters, provides a transparent and cost-effective option when wiring costs are prohibitively expensive.

PART 2—EXTERNAL WIRELESS COMMUNICATIONS

Summary

Just as mesh networking provides an excellent alternative to a wired solution at the sensor communication network level, advances in wireless technology allow the user to send data from the DAS to a remote server without the need for a hardwired connection. The use of GSM/GPRS (cell) modems in the DAS provides a very cost-effective means of communications that piggybacks on the structure built out for cell phone communications.

WIRED EXTERNAL SOLUTIONS

Fig. 5 shows a closer look at the upper part of Fig. 1 and provides some details about how data is typically sent from the DAS to a remote server for logging and reports. The DAS provides several options for communications,

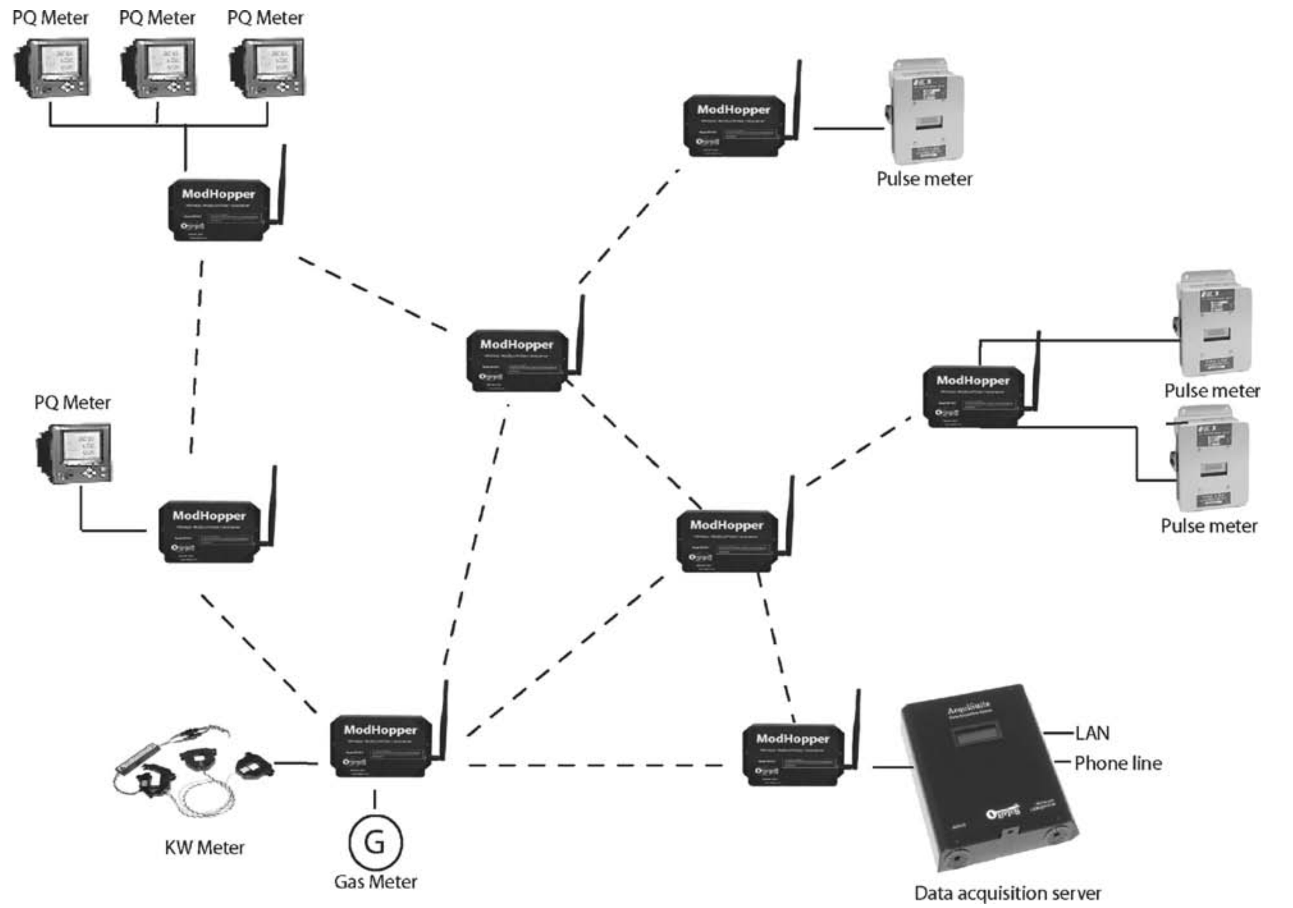


Fig. 4 Typical mesh network metering system.

depending on the needs of the user and available communications options. There are usually two alternatives for connecting to either a remote or a local server:

- *Existing LAN*—If the DAS is located in a facility that has an existing LAN, this provides an excellent mechanism for getting data from the DAS to a server. The DAS is IP-addressable, which means that a network-connected DAS can provide two-way communications and allow the user to see near real-time data in addition to interval information.
- *Phone line*—If there is no LAN connection (or if the IT department cannot or will not make a connection available), the DAS also provides the option to use an on-board modem to dial in or out to the system. If the user only needs to upload data (dial-out), the phone line can be shared with other devices such as fax machines.

Details of the file upload in a wired system are beyond the scope of this entry, but a typical upload session on an existing LAN would have the following elements:

1. The DAS initiates a connection to the remote server via the LAN by accessing a URL (e.g., http://www.obvius.com) where the server is located. The DAS provides a user name and password to the server, which uses this information to access the existing database records.
2. Once the connection is established and verified, the DAS will upload data collected since the last upload, in most cases using HTTP or FTP protocols.
3. Once the data is uploaded and verified, the DAS can optionally get additional information from the server, such as time checks.
4. Finally, the session is terminated.

The process for using a phone line connection is very similar, with the primary difference being that the upload speeds are slower (think dial-up modem vs DSL) and the connection is a PPP session.

WIRELESS OPTIONS FOR EXTERNAL COMMUNICATIONS

There are two basic applications for wireless communication from the DAS—one for inside the building to a LAN and one for totally wireless communications.

We will first look briefly at the simpler of the two, wireless communications within the building.

In this scenario, we are simply substituting a wireless access point for the wired LAN connection that then connects to another wireless access point on the local LAN. Because the DAS is an IP-addressable device like any workstation, there is no special setup required, and the functionality is exactly the same as that shown for the wired LAN connection in Figs. 5 and 6.

The second option for wireless external connectivity uses the existing cell phone networks built around the country to bypass any local connection. The primary reasons for doing this are the following:

- Lack of available LAN or phone lines
- Security concerns from IT personnel about using existing LANs
- High cost of acquiring and maintaining either phone or LAN connections
- Delays in getting phone or LAN connections installed
- Ease of installation

WHAT ADDITIONAL HARDWARE IS REQUIRED?

In the simplest terms, the only real change from the LAN to the phone-connected version is that a cell phone modem is used in place of the typical RJ11 modem. This modem is connected to an antenna mounted on the DAS, but these are the only functional differences between the cell modem version and any other DAS.

HOW DOES IT WORK?

In the olden days (think 1990s), cell phone communication was based on taking analog voice signals from one user and transmitting them to an analog receiver held by another user on the other end. Today, analog transmission is virtually nonexistent—ithas been replaced by digital cell technology (basically a process that converts the analog voice signal to digital for transmission that is decoded on the other end). The evolution from analog to digital is beneficial for data communication because data transmission on the Internet is inherently digital, and thus ideally suited for the digital networks (Fig. 7).

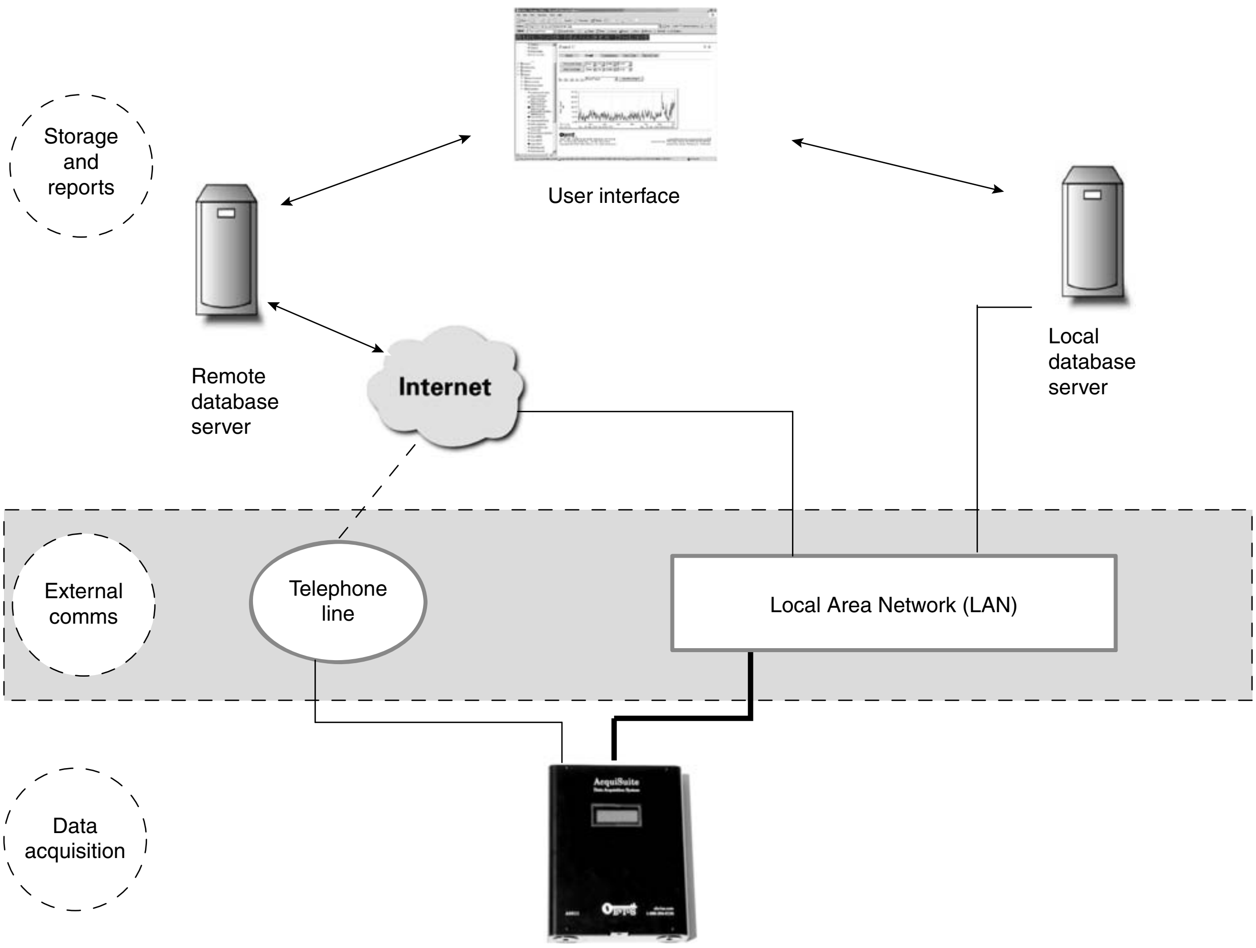


Fig. 5 Wired paths to the Web or a local database.

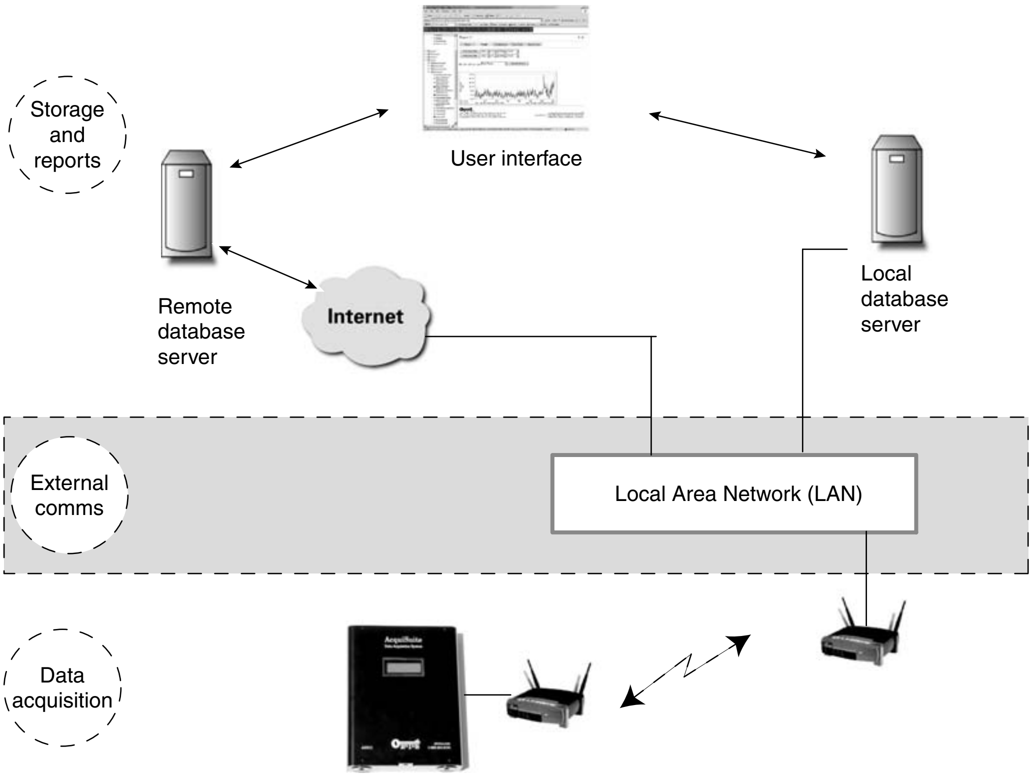


Fig. 6 Wireless Ethernet connection.

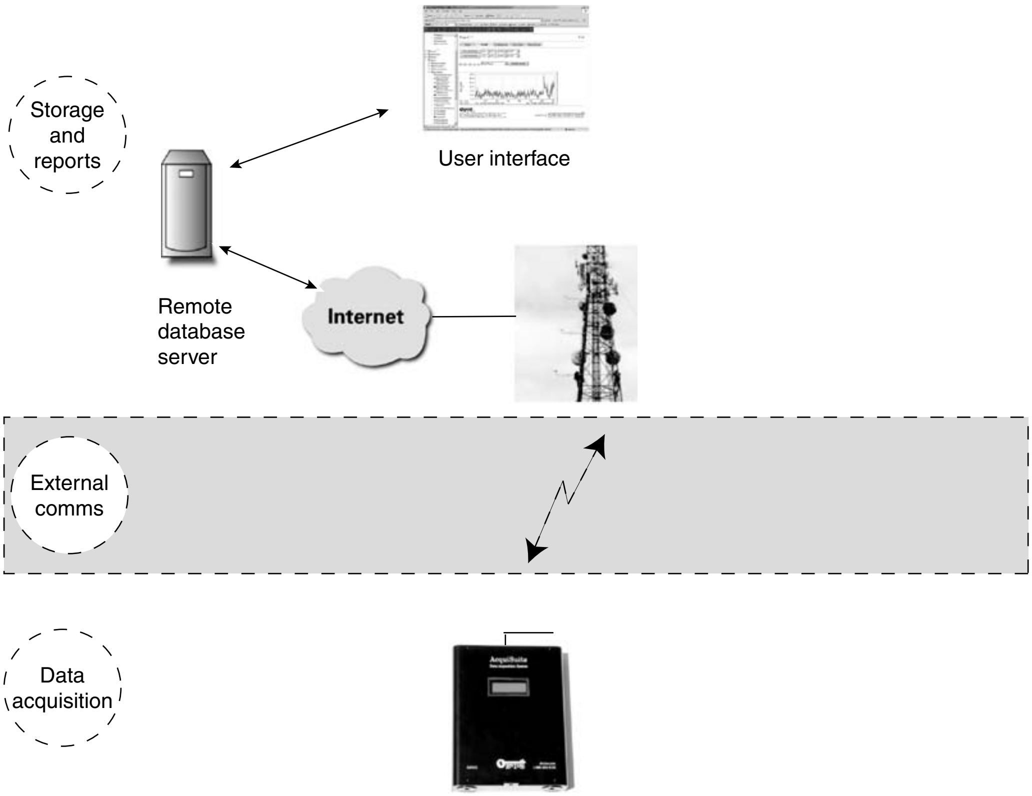


Fig. 7 Cell-based Web connection.

The DAS is functionally the same device used in the hardwired (LAN or phone modem) connection shown earlier in this entry. To use the cell phone system for data transmission, the phone modem module is replaced with a cell modem that connects to the cellular network just like any cell phone. The system relies on the use of the GSM network using GPRS to transmit data packets wirelessly over the cell network. This is the same system and service used by cell providers to allow users to view email and browse the internet using cell phones and PDAs.

Implementing this method of communication requires a SIM card (available from the wireless carrier) and a monthly data plan from the same carrier (e.g., Cingular) that typically runs $20–$30 per month for most DAS installations.

WHY USE THE CELL NETWORK INSTEAD OF A LAN OR PHONE LINE?

In addition to the ongoing monthly costs of cell service, the GSM version of the DAS is also typically 20%–30% more expensive than an identical LAN version, and the upload times are much slower. So, the obvious question is: Why anyone would use the cell system? There are a number of reasons why the GSM version may be preferable in some cases, despite the higher costs:

- *Costs of LAN or phone connections*—In many cases, it may be prohibitively expensive to add a network drop or phone line because of either the end user's policies or physical limitations (long wire runs, firewalls, etc.). In cases like this, the incremental costs of installing GSM may be less than the hardwired costs.
- *IT security concerns*—The LAN-based DAS is designed to function using the existing network without creating any security concerns. Because the GSM DAS has no connection to the existing network or phone system, any concerns are alleviated.
- *Installation delays*—Adding a network drop or phone line frequently involves coordinating with either other departments or contractors. In many installations, the lead times for getting the lines set and activated is the longest part of the installation; whereas the GSM system will be operating without the need for any outside resources the same day it is installed.

The most appealing aspects of the cell-based DAS are that it functions as a self-contained system without relying on existing networks or departments for installation or operation, and it can be up and sending data within minutes of the installation.

WHAT DOES THE FUTURE HOLD?

There are two primary forces driving improvements in the market for wireless submetering systems. First, advances in radio technology (both at the RS485 meter level and at the cellular system level) that are focused on broader markets such as consumer cell phones will be adopted by the energy information industry. Second, as the market for energy information grows, companies (such as Obvius) will emerge that specialize in the development of hardware and software specifically for gathering energy information, which will improve the functionality and ease of installation.

In general, it is reasonable to expect the following changes in the market:

- *Lower costs*—As companies outside the energy information industry (such as cell carriers) expand their offerings and increase the volume of sales, there will be lower costs for radio technologies and for the cellular service.
- *Easier installation*—The increased focus of specialized companies in this market will produce hardware and firmware that is focused on minimizing the time and costs associated with installation.
- *Broader coverage*—As the cellular companies expand and improve the coverage of their networks, it is reasonable to assume that some areas and locations that do not provide cell coverage today will be accessible in the future.

CONCLUSION

As we have seen, the development of application-specific wireless products designed to serve the submetering market has made the implementation of submetering networks much more cost effective. It is now practical to use wireless transceivers to transmit data within existing buildings or from building to building in a campus or base environment. In addition, the explosion of wireless Web options (including GPRS and WiFi) provides managers of submetering networks with wireless options that bypass the issues involved in accessing data via LANs or phone lines. For managers considering the installation of energy information systems, the cost of installing these systems is likely to drop even more in the future as more companies create new solutions and adapt existing technologies to data acquisition.

BIBLIOGRAPHY

1. Brambley, M.R.; Kintner-Meyer, M.; ONeill, P.J. Wireless sensor applications for building operation and management. In *Web Based Energy Information and Control Systems—Case Studies and Applications*; Capehart, B.L., Capehart, L.C., Eds., Fairmont Press: Lilburn, GA, 2005.
2. Standard for Information Technology, Telecommunications and Information Exchange Between Systems, Local and

Win–Wire

Metropolitan Area Networks, Specific Requirements, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher Speed Physical Layer Extension in the 2.4 GHz Band*. EE 802.11-1997. Institute of Electrical and Electronic Engineers: New York.

3. Katipamula, S.; Brambley, M.R.; Bauman, N.N.; Pratt, R.G. *Enhancing Building Operations Through Automated Diagnostics: Field Test Results*; Proceedings of the Third International Conference for Enhanced Building Operations, Texas A&M University, College Station: Texas, 2003.
4. Kitner-Meyer, M.; Brambley, M.R.; Carlon, T.A.; Bauman, N.N. Wireless Sensors: Technology and Cost-Savings for Commercial Buildings. *Information and Electronic Technologies; Promises and Pifalls*. In Teaming for Efficiency: Proceedings 2002 ACEEE Summer Study on Energy Efficiency in Buildings: Aug. 18–23, 2002; American Council for Energy Efficient Economy: Washington, DC, 2003; Vol. 7, pp. 121–134.
5. Kitner-Meyer, M.; Brambley, M.R. Pros & cons of wireless. ASHRAE J. **2002**, *44* (11), 54–61.
6. Kitner-Meyer, M.; Conant, R. *Opportunities of wireless sensors and controls for building operation*. 2004 ACEEE Summer Study on Energy Efficiency in Buildings. American Council for Energy Efficient Economy: Washington, DC, 2004.
7. Roundy, S.; Wright, P.K.; Rabaey, J.M. *Energy Scavenging for Wireless Sensor Networks with Special Focus on Vibrations*; Kluwer Academic Publishers: Boston, MA, 2003.
8. Su, W.; Akun, O.B.; Cayirici, E. Communication protocols for sensor networks. In *Wireless Sensor Networks*; Raghavendra, C.S., Sivalingam, K.M., Znati, T., Eds., Kluwer Academic Publishers: Boston, MA, 2004; 21–50.
9. Ye, W.; Heideman, J. Medium access control in wireless sensor networks. In *Wireless Sensor Networks*; Raghavendra, C.S., Sivalingam, K.M., Znati, T., Eds., Academic Publishers: Boston, MA, 2004; 73–91.
10. Michael, R.B.; Kintner-Meyer, M.; ONeill, P.J. Wireless sensor applications for building operation and management. In *Web Based Energy Information and Control Systems-Case Studies and Applications*; Capehart, B.L., Capehart, L.C., Eds., Fairmont Press: Lilburn, GA, 2005.

Wireless Applications: Mobile Thermostat Climate Control and Energy Conservation

Alexander L. Burd

Galina S. Burd

Advanced Research Technology, Suffield, Connecticut, U.S.A.

Abstract

The annual energy savings (AES) associated with the application of the wireless mobile thermostat for an average two-story residential house in the United States are projected to be approximately in the range of 4.5%–15.7% for heating and 3.1%–8.8% for cooling, respectively, depending on geographic location. The simple payback for the new control system serving space heating and space cooling loads is estimated to be approximately 1.9–3.5 years.

TYPICAL CLIMATE CONTROL ARRANGEMENT IN RESIDENTIAL BUILDINGS

The so-called forced air system with central furnace is the most frequently used system in the United States. The forced air system is utilized for space heating in 56.1 million households out of 101.5 million residential houses (or in 55.3% houses), including apartment complexes. The forced air system is quite extensively used for air conditioning (space cooling) and heating, utilizing the same ductwork to distribute heating and cooling throughout a house. Approximately 47.5 million houses out of 73.7 million single-family houses (or 64.5% houses) equipped with central air conditioning system also utilize a central furnace.[1]

About 30.1 million households (or 40.8%) among 73.7 million single-family houses with space heating and cooling are either two-story (26.9 million or 36.5%) or three-story (3.2 million or 4.3%) houses. The rest of the houses (43.6 million) are single-story houses (about 59.2%). Of these 73.7 million households, only approximately 400,000 households (about 0.55%) utilize two zones, with each zone (most frequently subdivided by floor) served by a dedicated furnace, which has its own stationary thermostat to control heating and cooling.[2]

A vast majority of residential houses with forced air systems have only one thermostat to control both heating and cooling modes of operation.

PREVIOUS AND CURRENT INVESTIGATIONS

The previous conducted investigation established that a significant temperature differential exists between the first and second floors when control of space heating and cooling is implemented from the central stationary thermostat located in the living room on the first floor. The air temperature is also subject to significant variation within each floor.[4] A temperature differential between various locations in the rooms is a function of many factors, such as the distance to the heating/cooling discharge air outlets and the windows, as well as occupancy level, heat gains from appliances, wind direction, orientation toward the sun, etc.

The application of a mobile thermostat allows for control of air temperature on demand at any time in any location of the building. This leads to better indoor climate control and reduces annual thermal and electrical energy consumption.[4,5] The potential AES for an average household in New England were analyzed earlier by Burd and Burd.[6] This paper evaluates the potential AES that could be realized via utilization of the new control system, which features a mobile thermostat in residential buildings at different geographic locations in the United States. Although the application of a mobile thermostat would reduce energy consumption in residential buildings due to multiple factors (including horizontal temperature variation within each floor), in our feasibility evaluation for the developed system, we conservatively considered energy savings that are caused by the vertical temperature variation only. These savings are associated with the temperature differential due to forced and natural convections[3] between floors in multistory buildings (i.e., two- and three-story houses).

INVESTIGATION OF TEMPERATURE DIFFERENTIAL BETWEEN SECOND AND FIRST FLOOR IN A RESIDENTIAL TOWNHOUSE

This investigation was conducted in a typical residential townhouse built in 1986 and located in the state of

Keywords: Residential; Comercial; Wireless mobile thermostat; Space heating; Air conditioning; Control system; Energy conservation.

Connecticut (New England region of the United States). The two-floor townhouse has approximately 92.9 m^2 (1000 ft^2) of total area. The height of each floor is 2.44 m (8 ft). The townhouse has a living room, a kitchen, and a half bathroom located on the first floor and two bedrooms and a full bathroom on the second floor (the smaller bedroom is used as a study room). In addition, the townhouse has a full unfinished basement used as storage. The basement is also heated and cooled via two diffusers supplying the warm and cold air; the forced-air furnace and associated air distribution ductwork are located in the basement. The townhouse is situated in the middle of a four-unit residential building and has easterly- and westerly-oriented outside exposures.

The monitored residential townhouse was built after the new ASHRAE ventilation norms were adapted and, thus, could be rated as a "tight" house. The infiltration rates in ACH (air change per hour) for residential buildings in the United States were adapted from the data published by McQuiston. The ACH indicates the ratio of the hourly infiltration air volume to the volume of the house. These rates vary from 0.51 ACH ("tight" houses) to 1.3 ACH ("loose" houses). Natural gas and electricity are used for space heating and space cooling, respectively. The furnace has an evaporator (cooling coil); the cooling compressor and condenser units of the cooling system are located outside of the house. The air treated in the furnace (heated or cooled) is delivered throughout the house via ductwork, supply diffusers, and registers, and then it is returned back to the furnace via return grills and ductwork. Two people live in the townhouse. The house has one stationary, nonprogrammable thermostat located in the living room, which controls both heating and cooling.

The temperature differential between the second and the first floors (TDSFF) depends on both forced and natural convection. For the purpose of this analysis, in Fig. 1, we considered a separate impact on TDSFF from the forced and natural air convection.

Natural convection is present in one way or the other during the entire heating season. However, its impact is more significant when the forced air space heating system is turned off. Because of that, natural convection impact is more pronounced during relatively warm outdoor air temperatures with fewer hours of space heating operation.

As opposed to natural convection, the impact of forced air convection is more noticeable during lower outdoor air temperatures with frequent space heating system operation. Therefore, the TDSFF, due to the combined effect of the forced air and natural convection, will be somewhat higher during lower outdoor air temperatures, as compared to the high outdoor air temperatures (see Fig. 1).

The combined impact of natural and forced air convection on TDSFF for the entire heating season is represented by line 3 in Fig. 1. The data in line 3 represent the summation of the data in line 1 and line 2.

According to Fig. 1, the average TDSFF during heating season, with an outdoor air temperature of 4.7°C (40.5°F), will be close to about 2.2°C (36°F).

The stationary thermostat set point in the townhouse (for the monitored data used in Fig. 2) was maintained at about 18°C (64.4°F).

WIRELESS MOBILE THERMOSTAT CONTROL SYSTEM

The accuracy of maintaining a desired air temperature in the building greatly depends on the utilized control system.

The closer the control device is to the area where the targeted temperature value is critical to maintain, the more accurate the control becomes. Obviously, the best and most precise control could be achieved by combining various stages of control and, eventually, by utilizing a final (ultimate) control stage when the control device is

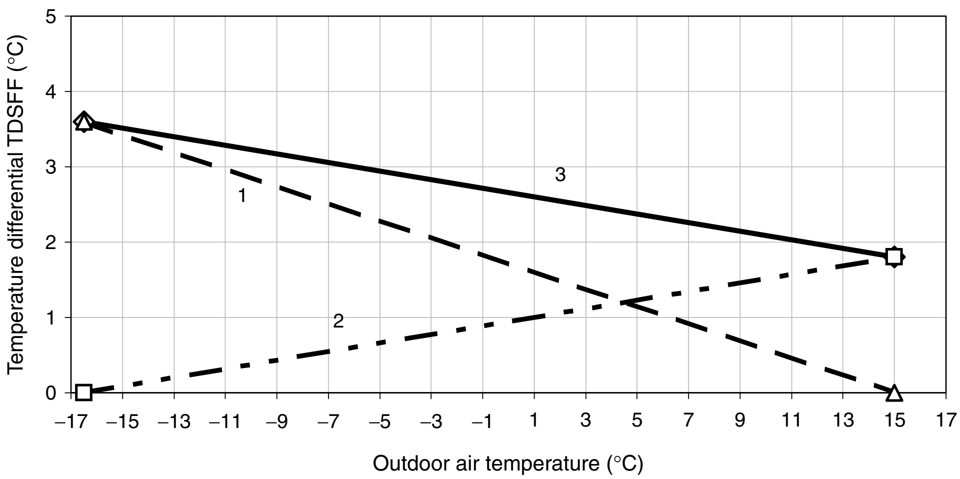


Fig. 1 Variations in outdoor air temperature and temperature differential between second and first floors (TDSFF). 1. Forced air convection impact; 2. Natural air convection impact; 3. Combined impact of forced air and natural air convection.

Win–Wire

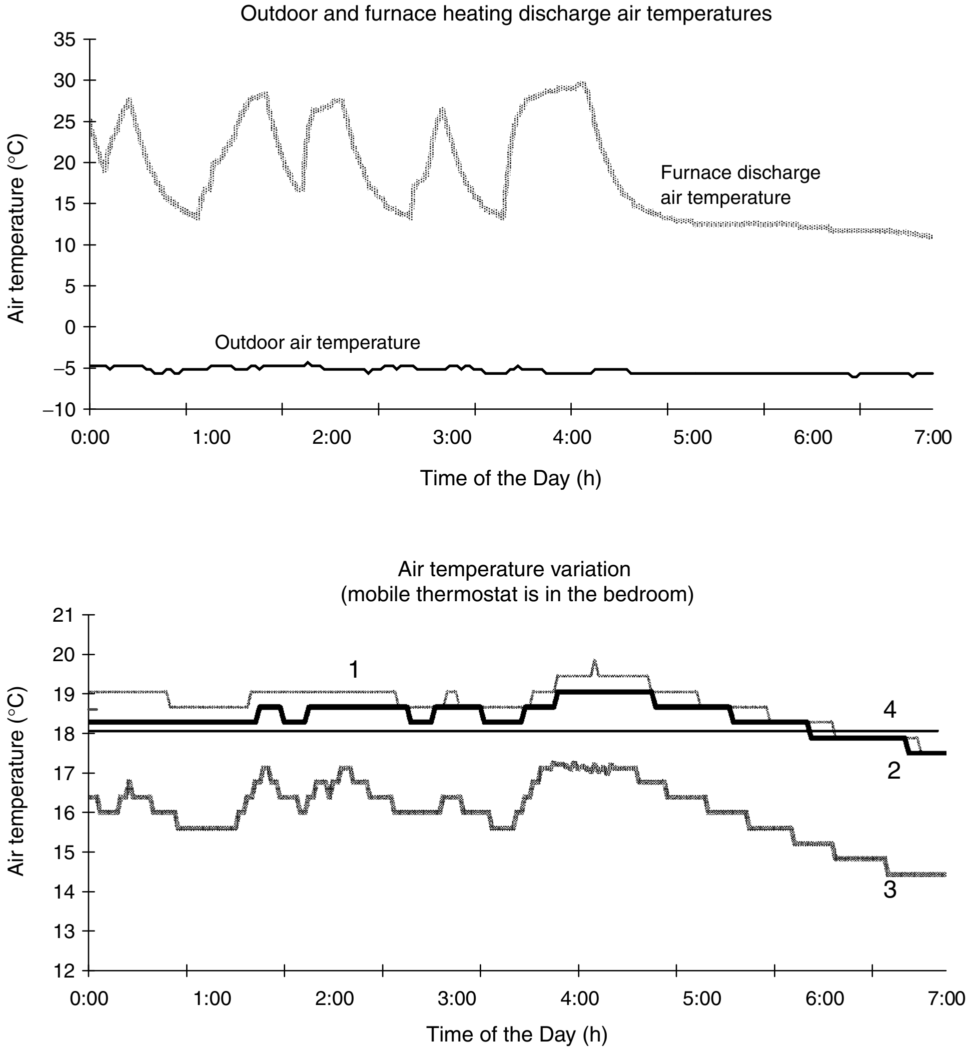


Fig. 2 Mobile thermostat operation during nighttime (heating mode). 1. Study room; 2. Bedroom; 3. Living room (dinner table); 4. Thermostat temperature set point.

located at the place where a desired air temperature value has to be maintained.

An innovative control system that utilizes a wireless mobile thermostat has been developed.[7] The mobile thermostat is a battery-operated device, which facilitates the implementation of the final stage of control to accurately balance the heat loss and heat supply in order to maintain the required air temperature in a particular area of the house. A prototype of the wireless mobile thermostat was manufactured to conduct testing of the thermostat's performance.

The prototype consists of two parts: a receiver and a transmitter. The prototype was installed in the investigated townhouse. The mobile thermostat's transmitter (remote temperature sensing element) was moved to different areas throughout the house (such as the bedroom, living room, study, etc.) to control air temperature as necessary.

The transmitter of the wireless mobile thermostat is capable of sending a remote radio signal to the receiver which, in turn, activates or deactivates the furnace's heating or cooling systems to maintain a set point temperature at the thermostat's transmitter location. This remote signal is proportional to the temperature differential between the wireless mobile thermostat's set point temperature and the air temperature at the thermostat's location. This proportionality is realized via the length of time during which the signal is sent from the thermostat to the receiver. The higher the initial deviation of the air temperature at the thermostat's location from the set point, the longer the time period during which the signal is transmitted from the transmitter to the receiver. This, in turn, would increase the operating time for the heating/cooling unit to heat/cool the area. A reverse procedure would ocurr if the initial temperature at the thermostat's location was closer to the thermostat's set point. Another distinctive feature of the wireless mobile thermostat is its low inertia and ability to react quickly to any temperature changes at the thermostat's location. A detailed presentation of the wireless mobile thermostat system's major elements is given in Burd and Burd.[4]

WIRELESS MOBILE THERMOSTAT OPERATION DURING HEATING MODE

The results of the wireless mobile thermostat's operational tests are shown in Fig. 2.

The upper part of the graph in Fig. 2 shows that at nighttime, the outdoor air temperature was near −6.5°C (20.3°F). The maximum value of the furnace's discharge air temperature measured in the basement was near 30°C (86°F). The furnace cycles on and off.

The mobile thermostat was located on the nightstand in the bedroom, and its temperature set point was 18°C (64.4°F). The lower part of the graph shows that the temperature in the bedroom was maintained close to the thermostat's set point [maximum deviation did not exceed +1°C (1.8°F) and −0.5°C (0.9°F)]. The mobile thermostat overrode the stationary thermostat control and turned the furnace on and off, as necessary (as shown in the upper portion of the graph), to maintain the required temperature at the mobile thermostat's transmitter location.

While the temperature in the bedroom (as well as in the study) closely followed the wireless mobile thermostat temperature set point, the temperature in the living room was maintained at a substantially lower magnitude.

The air temperature difference between the bedroom and the study room was not quite noticeable. On the other hand, the air temperature in the living room was significantly lower—by 1°C (1.8°F) to 4°C (7.2°F)—than the temperature at the mobile thermostat location. This demonstrates the ability of the wireless mobile thermostat to save energy for space heating by maintaining the desired temperature in the occupied rooms upstairs on demand, while keeping a lower temperature on the first floor when it is not occupied.

PROJECTED ANNUAL ENERGY SAVINGS IN RESIDENTIAL BUILDINGS

The projected AES due to the application of the wireless mobile thermostat are shown in Table 1. The energy savings were calculated for a number of the selected U.S. cities, which represent a wide variety of climatic conditions.

The design space heating outdoor air temperature conditions for these cities vary from −29.4°C (−20.9°F) for Bismark (North Dakota) to 7.8°C (46°F) for Miami (Florida). Table 1 indicates that the average (per heating season) outdoor air temperature for the selected cities vary from −1.1°C (30.1°F) for Bismarck (North Dakota) to 14°C (57.2°F) for Miami (Florida). The annual space heating run time for these cities (when the outdoor temperature is lower than the space heating balance temperature) varies from 894 h for Miami (Florida) to 7815 for Seattle (Washington).

The design space's cooling outdoor air temperature conditions differ from 43.3°C (109.9°F) for Phoenix (Arizona) to 29.4°C (84.9°F) for Seattle (Washington). Table 1 also demonstrates that the average outdoor air temperature over the cooling season for the selected cities varies from 26.1°C (79°F) for Portland (Maine) to 31.3°C (88.3°F) for Phoenix (Arizona). The annual space cooling run time for these cities (when the outdoor air temperature is higher than the space cooling balance temperature) varies from 237 h for Seattle (Washington) to 5390 h for Miami (Florida).

The cumulative annual operating time for space heating and cooling ranges from the maximum of 8052 h (91.9% of the length of the entire year) for Seattle (Washington) to the minimum of 6141 h (70.1% of the length of the entire year) for Tampa (Florida).

The application of the mobile thermostat would allow a user to setback the air temperature during the heating period and to setforward the air temperature during a cooling period. We used a simplified engineering method of calculation based on the assumption that the energy consumption for space heating and space cooling can be expressed as a linear function of the temperture differential between the two major parameters, averaged over a heating/cooling season. These parameters are: indoor dry-bulb air temperature set point and oudoor dry-bulb air temperature. This approach assumes that the outdoor dry-bulb air temperature can be used as a defining parameter, impacting energy conservation. This is utilized for the purpose of initial evaluating analysis only. A detailed presentation of the methodology for energy savings calculations is given in Burd and Burd.[6]

Table 1 shows the potential energy savings associated with the utilization of the mobile thermostat. The AES due to lowering or increasing a stationary thermostat's set point temperature by +1°C during the heating or cooling mode of operation were calculated by the formula:

$$\mathrm{AES} = \frac{\Delta \mathrm{T}}{\pm(\mathrm{HCS} - \mathrm{T}_{\mathrm{AV.OUT.HC}})} \times 100,\ \%$$

where ΔT, the magnitude of lowering or increasing a set point temperature with the mobile thermostat, °C (°F); HCS, the current heating or cooling set point temperature maintained by the stationary thermostat, °C (°F); and $T_{AV.OUT.HC}$, the average outdoor air temperature during the heating or cooling season, respectively, °C (°F).

Table 1 denotes that the potential energy savings associated with the 1°C (1.8°F) setback in air temperature via the wireless mobile thermostat during heating season would vary from 4.7% (for Bismarck, North Dakota) to 16.7% (for Miami, Florida). The potential energy savings associated with the 1°C (1.8°F) setforward in air

Table 1 Potential annual energy savings due to mobile thermostat application for space heating and space cooling for selected cities in U.S.A.

No	City	State	North latitude (Degrees)	Design outdoor air temperature for heating (°C)	Average per heating season outdoor air temperature (°C)	Annual space heating system operating time (h)	Potential energy savings for space heating per 1°C thermostat temperature setback (%)	Design outdoor air temperature for cooling (°C)	Average per cooling season outdoor air temperature (°C)	Annual space cooling system operating time (h)	Potential energy savings for space cooling per 1°C thermostat setforward (%)	Annual space heating and space cooling systems operating time (h)	Cumulative percentage of space heating and space cooling systems operating time during a year (%)
1	Bismarck	North Dakota (ND)	47	−29.4	−1.1	6,896	4.7	33.9	27.8	824	25.4	7,720	88.1
2	Chicago	Illinois (IL)	42	−21.1	4.4	6,075	6.4	32.8	27.6	1,193	27.3	7,268	83.0
3	Portland	Maine (ME)	46	−5.6	4.5	7,265	6.5	32.2	26.1	383	45.0	7,648	87.3
4	Buffalo	New York (NY)	43	−16.7	4.5	6,704	6.5	30.0	26.3	616	40.9	7,320	83.6
5	Hartford	Connecticut (CT)	42	−16.7	4.7	6,455	6.5	32.8	27.2	934	30.0	7,389	84.3
6	Baltimore	Maryland (MD)	39	−11.7	6.2	5,717	7.2	33.9	27.5	1,308	27.7	7,025	80.2
7	Charleston	South Carolina (SC)	33	−3.9	6.8	5,703	7.6	34.4	27.2	1,206	30.0	6,909	78.9
8	Seattle	Washington (WA)	47	−5.0	9.1	7,815	9.1	29.4	26.3	237	40.9	8,052	91.9
9	Dallas	Texas (TX)	33	−8.3	9.7	4,079	9.7	37.8	28.6	3,029	21.4	7,108	81.1
10	Phoenix	Arizona (AZ)	34	1.1	11.6	3,305	11.8	43.3	30.2	3,861	15.8	7,166	81.8
11	Tampa	Florida (FL)	28	2.2	13.1	2,262	14.4	33.3	27.4	3,879	28.6	6,141	70.1
12	Miami	Florida (FL)	26	7.8	14.0	894	16.7	32.8	26.9	5,390	32.7	6,284	71.7

Assumptions:1. Space heating balance temperature is assumed to be 16.7°C (62°F)—below 62°F heating is required;2. Space cooling balance temperature is assumed to be 23.9°C (75°F)—above 75°F cooling is required;3. Space heating and space cooling design dry-bulb air temperatures are assumed at their 99.6 percentile.

No	City	State	Annual space heating energy consumption (kWh)	Annual space heating energy savings per 1°C temperature setback (kWh)	Annual space cooling energy consumption (kWh)	Annual space cooling energy savings per 1°C thermostat temperature setforward (kWh)	Cumulative annual space heating and space cooling energy consumption (kWh)	Percentage of space heating energy consumption (%)	Percentage of space cooling energy consumption (%)	Annual space heating and cooling savings per 1°C temp. setback for heating and setforward for cooling (kWh)	Cumulative annual relative savings percentage for space heating and space cooling per 1°C temp. setback and setforward (%)
1	Bismarck	North Dakota (ND)	35,350	1,679	459	116	35,809	98.7	1.3	1,795	5.0
2	Chicago	Illinois (IL)	23,007	1,479	617	168	23,624	97.4	2.6	1,647	7.0
3	Portland	Maine (ME)	27,415	1,769	120	54	27,536	99.6	0.4	1,823	6.6
4	Buffalo	New York (NY)	25,298	1,632	212	87	25,511	99.2	0.8	1,719	6.7
5	Hartford	Connecticut (CT)	24,010	1,572	439	132	24,449	98.2	1.8	1,703	7.0
6	Baltimore	Maryland (MD)	19,254	1,392	666	185	19,920	96.7	3.3	1,576	7.9
7	Charleston	South Carolina (SC)	18,358	1,388	567	170	18,926	97.0	3.0	1,559	8.2
8	Seattle	Washington (WA)	20,823	1,903	82	33	20,905	99.6	0.4	1,936	9.3
9	Dallas	Texas (TX)	10,207	993	1,994	427	12,201	83.7	16.3	1,420	11.6
10	Phoenix	Arizona (AZ)	6,795	805	3,450	545	10,244	66.3	33.7	1,349	13.2
11	Tampa	Florida (FL)	3,824	551	1,915	547	5,740	66.6	33.4	1,098	19.1
12	Miami	Florida (FL)	1,306	218	2,323	760	3,629	36.0	64.0	978	26.9

Assumptions:1. Space heating and space cooling energy consumption changes in direct proportion to the difference between indoor and outdoor air temperatures;2. Historical energy consumption data for the monitored two-story townhouse in Hartford, CT was assumed to be a base for the calculations 0.000293 conversion factor from Btu to kWh.

temperature via the wireless mobile thermostat during cooling season would vary from 15.8% (for Phoenix, Arizona) to 45% (for Portland, Maine). The projected relative value of energy savings [for 1°C (1.8°F) reset in temperature set point] during the cooling mode of operation is higher than for the heating mode of operation because the temperature differential between the thermostat set point and average seasonal outdoor air temperature is lower for cooling as compared to heating.

The continuation of Table 1 shows annual heating and cooling energy consumption as well as cumulative annual energy consumption for space heating and cooling for the considered cities. The heating and cooling annual energy consumption for the monitored townhouse in Connecticut was calculated based on the actual electrical and gas meters data. The annual heating energy consumption (HEC) for the houses in the selected cities was calculated by the formula:

$$HEC = \frac{HEC_{MH}(T_{INH} - T_{AV.OUT.HSC})}{T_{IN.H} - T_{AV.OUT.MH}} \times \frac{ASHOT_{SC}}{ASHOT_{MH}}, \text{ kWh/yr},$$

where HEC_{MH}, annual energy consumption for heating in the monitored house, kWh (MJ); $T_{IN.H}$, the air temperature inside the house during heating season was assumed to be 20°C (68°F); $T_{AV.OUT.HSC}$, average (per heating season) outdoor air temperature for the selected cities, °C (°F); $T_{AV.OUT.MH}$, average (per heating season) outdoor air temperature for the monitored house, °C (°F); $ASHOT_{SC}$, annual space heating operating time for the selected cities, h; and $ASHOT_{MH}$, annual space heating operating time for the monitored house, h.

The annual cooling energy consumption (CEC) for the houses in the selected cities was calculated by the formula:

$$CEC = \frac{CEC_{MH}(T_{AV.OUT.CSC} - T_{IN.C})}{T_{AV.OUT.MH} - T_{IN.C}} \times \frac{ASCOT_{SC}}{ASCOT_{MH}}, \text{ kWh/yr},$$

where $T_{IN.C}$, the air temperature inside the house during cooling season was assumed to be 23.9°C (75°F); CEC_{MH}, the annual energy consumption for cooling in the monitored house, kWh (MJ); $T_{AV.OUT.CSC}$, the average (per cooling season) outdoor air temperature for the selected cities, °C (°F); $T_{AV.OUT.MH}$, the average (per cooling season) outdoor air temperature for the monitored house, °C (°F); $ASCOT_{SC}$, the annual space cooling operating time for the selected cities, h; and $ASCOT_{MH}$, the annual space cooling operating time for the monitored house, h.

Table 1 (continuation) also indicates the cumulative potential energy savings of a 1°C (1.8°F) temperature setback for space heating and temperature setforward for cooling. For the majority of the selected cities, the HEC far exceeds the cooling energy consumption. This is only not the case for Miami (Florida), where the annual energy consumption for cooling (64%) is higher than for heating (36%).

Table 2 demonstrates that the cumulative annual relative energy savings for heating and cooling with the wireless mobile thermostat vary from 5.0% for Bismark (North Dakota) to 26.9% for Miami (Florida) of the current (baseline) energy consumption for the considered buildings for each °C (1.8°F) of the temperature setback and setforward.

Table 2 illustrates the potential annual energy cost savings per 1°C (1.8°F) in temperature setback and setforward for space heating and space cooling, respectively, for the geographic locations considered in Table 1. The cumulative annual space heating and space cooling energy cost savings range from approximately $102 for Portland (Maine) to $447 for Phoenix (Arizona).

Table 3 shows potential annual energy cost savings due to the wireless mobile thermostat application, considering the results of temperature monitoring and the occupancy schedule in the representative house.[6] Based on the results of temperature monitoring discussed earlier, we assumed that the utilization of the wireless mobile thermostat would allow reduction of the air temperature in the house by 2.2°C (3.96°F) during the heating season, while the upstairs areas are occupied. Based on the results of temperature monitoring,[5,6] we also assumed that the use of the wireless mobile thermostat would allow the user to increase the air temperature of the house by 0.5°C (0.9°F) during the cooling season when it is occupied and occupants are on the second floor. The occupancy schedule assumes that the daily, average, per-week occupied, and nonoccupied time is 14.4 and 9.6 h, respectively. The occupancy schedule also assumes that the occupants spend 30 and 70% of the occupied time in the house downstairs and upstairs, respectively.

Considering the above occupancy schedule, the overall wireless mobile thermostat's daily setback for space heating and setforward for space cooling is assumed to be approximately 0.9°C (1.62°F) and 0.2°C (0.36°F), respectively. These additional savings are once again projected, compared to the existing stationary thermostat located on the first floor of the two-story house. The cumulative annual space heating and space cooling savings vary from $64 for Miami (Florida) to $117 for Phoenix (Arizona). The simple payback for the wireless mobile thermostat would vary from 1.9 years for Phoenix (Arizona) to 3.5 years for Miami (Florida). For the majority of geographic locations considered in the study, the mobile thermostat would have a simple payback of 2.4–2.7 years.

An application of the wireless mobile thermostat in residential buildings with only space heating or space cooling loads might be less advantageous from an

Table 2 Potential annual energy cost savings due to mobile thermostat application per one °C temperature setback for heating and temperature setforward for cooling

No	City	State	Annual space energy heating savings (kWh)	Annual space heating energy cost savings ($)	Annual space cooling energy savings (kWh)	Annual space cooling energy cost savings ($)	Cumulative annual space heating and cooling energy cost savings ($)
1	Bismarck	North Dakota (ND)	1,679	84	459	54	138
2	Chicago	Illinois (IL)	1,479	74	617	73	147
3	Portland	Maine (ME)	1,769	88	120	14	102
4	Buffalo	New York (NY)	1,632	81	212	25	106
5	Hartford	Connecticut (CT)	1,572	78	439	52	130
6	Baltimore	Maryland (MD)	1,392	69	666	79	148
7	Charleston	South Carolina (SC)	1,388	69	567	67	136
8	Seattle	Washington (WA)	1,903	95	82	10	105
9	Dallas	Texas (TX)	993	50	1,994	235	285
10	Phoenix	Arizona (AZ)	805	40	3,450	407	447
11	Tampa	Florida (FL)	551	27	1,915	226	253
12	Miami	Florida (FL)	218	11	2,323	274	285

Notes:1. The above calculations are conducted for 1°C in temperature setback and setforward for space heating and cooling, respectively;2. The installed cost of the mobile thermostat for the residential application is assumed to be $220 [10–12];3. Cost of natural gas used for space heating was assumed to be $1.46 for 29.3 kWh (100,000 Btu) or $0.0498/kWh;4. Cost of electricity used for space cooling was assumed to be $0.118/kWh.

Table 3 Potential annual energy cost savings due to mobile thermostat application considering occupancy pattern in the representative house

No	City	State	Potential energy savings for space heating thermostat temperature setback (%)	Annual space heating energy savings (kWh)	Annual space heating energy cost savings ($)	Potential energy savings for space cooling thermostat temperature setforward (%)	Annual space cooling energy savings (kWh)	Annual space cooling energy cost savings ($)	Cumulative annual space heating and cooling energy cost savings ($)	Simple payback period (yrs)
1	Bismarck	North Dakota (ND)	4.5	1,586	79	4.9	89	11	90	2.5
2	Chicago	Illinois (IL)	6.1	1,397	70	5.3	120	14	84	2.6
3	Portland	Maine (ME)	6.1	1,670	83	8.8	23	3	86	2.6
4	Buffalo	New York (NY)	6.1	1,541	77	8.0	41	5	82	2.7
5	Hartford	Connecticut (CT)	6.2	1,484	74	5.8	85	10	84	2.6
6	Baltimore	Maryland (MD)	6.8	1,315	66	5.4	130	15	81	2.7
7	Charleston	South Carolina (SC)	7.1	1,311	65	5.8	110	13	78	2.8
8	Seattle	Washington (WA)	8.6	1,797	90	8.0	16	2	91	2.4
9	Dallas	Texas (TX)	9.2	938	47	4.2	388	46	93	2.4
10	Phoenix	Arizona (AZ)	11.2	760	38	3.1	671	79	117	1.9
11	Tampa	Florida (FL)	13.6	520	26	5.6	372	44	70	3.1
12	Miami	Florida (FL)	15.7	206	10	6.4	452	53	64	3.5

Notes:1. Assumed average per week temperature setback for heating 0.94°C, 1.7°F;2. Assumed average per week temperature setforward for cooling 0.19°C, 0.35°F.

economical point of view, as compared to the houses with both loads.

Energy savings for the wireless mobile thermostat will be achieved due to minimization of what we called a "comfort satisfaction safety factor" (CSSF)—when the set point temperature at the stationary thermostat is maintained at a higher level during the heating season and a lower level during the cooling season to compensate for any deviations of the temperature at the occupants' location to ensure their satisfaction with the indoor climate control based on the emperical anticipation. Similarly, the occupant applies the CSSF for the first floor stationary thermostat set point to satisfy the required comfort conditions on the second floor.

The above leads to a logical conclusion that the utilization of the mobile thermostat would always produce energy savings as compared to a baseline energy consumption with a stationary thermostat. Obviously, the CSSF would be different for various users and applications.

The projected savings in residential buildings with a single stationary thermostat serving radiation or convector steam or hot water heating systems with natural air convection may be somewhat lower as compared to the forced air heating systems. This could be due to the reduced temperature differential between the second and first floors for systems with natural air convection. Further investigation would be necessary to verify this assumption.

In addition to individual residential houses, the developed control system could also be used in a variety of heating and cooling applications in commercial buildings, as well. Wireless mobile thermostat could also be instrumental in reducing energy consumption in district energy systems, such as district heating/cooling systems,[8] as well as for various industrial applications—for climate controls in zones served by rooftop units, where manufacturing equipment is frequently moved around the facility,[9] etc.

CONCLUSION

The wireless mobile thermostat—the ultimate stage of climate control—would allow a user to setback and setforward temperatures in the house for space heating and space cooling, respectively. The conducted initial investigation utilizing a simplified model for different geographic locations in the United States showed that the application of a wireless mobile thermostat in the two-story townhouse would allow savings of 4.5%–15.7% of annual space heating energy. Utilization of the wireless mobile thermostat would also allow savings of 3.1%–8.8% of energy for space cooling. The simple payback for the new control system serving space heating and space cooling loads would vary from approximately 1.9–3.5 years.

The mass application of the wireless mobile thermostat, considering the scale of energy consumption in residential buildings, which consume more than 20% of the total energy use—including residential, commercial, agricultural and transportation sectors—could be an important step in energy resource.[10]

The ever-increasing cost of fuel could become a contributing factor for wireless mobile thermostat utilization. In addition, the application of the wireless mobile thermostat will have a positive environmental impact due to emissions reduction at power plants as well as at residential heat generation systems.

REFERENCES

1. Energy Information Administration/Household Energy Consumption and Expenditures, U.S.A., 1997.
2. Energy Information Administration/Annual Energy Review, U.S.A., 1999.
3. ASHRAE Handbook *Fundamentals*; American Society of Heating, Refrigerating, and Air Conditioning Engineers, Inc.: Atlanta, GA, 1997.
4. Burd, A.L.; Burd, G.S. Mobile thermostat for climate control in residential buildings, Proceedings of CIBSE/ASHRAE Joint Conference 2000, paper H-1562, Conference Papers on CD ROM: 1–22, Dublin, Ireland, 2000.
5. Burd, A.L.; Burd, G.S. *Wireless Mobile Thermostat for Climate Control Enhancement in Buildings*. Proceedings of Healthy Buildings 2000, Espoo, Finland, 2000; Vol. 2, 763–768.
6. Burd, A.L.; Burd, G.S. Application of Mobile Wireless Thermostat and Energy Conservation and Comfort in Residential Houses. ASHRAE Transactions, Honolulu, U.S.A., 2002; Vol. 108, Part 2, 202–211.
7. Burd, A.L. Mobile Thermostat to Control Space Temperature in the Building, Patent Number 5, 419, 489, 1–6, U.S.A., 1995.
8. Burd, A.L. Deferred heat supply for space heating using a capacity-limiting device—a beneficial approach for district heating. ASHRAE Transactions: Research, Boston, U.S.A., 1997; Vol. 103, Part 2, 23–31.
9. Burd, A.L.; Burd, G.S.; De Maio, M. Smith and Wesson: The Story of a Chilled-Water Retrofit. HPAC Engineering, March 2005; Part. 1, 38–47 and July 2005, Part. 2, 24–37.
10. Wisconsin Energy Statistics. Wisconsin Energy Bureau, Department of Admnistration, U.S.A., 1997.

BIBLIOGRAPHY

1. An idea whose time has come…the wireless thermostat! Totaline Star. Replacement Component Division, Carrier Corporation; U.S.A., 1999; 1.
2. ASHRAE Handbook *Systems and Equipment*; American Society of Heating, Refrigerating, and Air Conditioning Engineers, Inc.: Atlanta, GA, 2000.
3. McQuiston, F.C. A study and review of existing data to develop a standard methodology or residential heating and cooling load calculations. ASHRAE Trans., *90* (2A), 102–135.
4. Wireless thermostat. RCI Automation, Specification, 2002.
5. Wireless thermostat. Southern California Air Conditioning Distributors. Specifications, 2002.